DOCUMENT RESUME

ED 431 015 TM 029 847

AUTHOR Margerum-Leys, Jon; Kupperman, Jeff; Boyle-Heimann, Kristen
TITLE Analytical and Methodological Issues in the Use of Qualitative Data Analysis Software: A Description of Three Studies.
SPONS AGENCY Spencer Foundation, Chicago, IL.
PUB DATE 1999-04-22
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Coding; *Computer Software; *Data Analysis; *Qualitative Research; *Research Methodology
IDENTIFIERS *NUDIST Computer Program

ABSTRACT

This paper presents perspectives on the use of data analysis software in the process of qualitative research. These perspectives were gained in the conduct of three qualitative research studies that differed in theoretical frames, areas of interests, and scope. Their common use of a particular data analysis software package allows the exploration of issues related to use of the software, QSR NUDIST (NUDIST). NUDIST combines the accessibility to text of a word processing program with the data handling capacity of a database program, and some reporting features of a spreadsheet. The first study examined the beliefs of nine teacher education students. The second study examined the effects on middle school students (n=14) of Internet use when the technology was provided in the home. The third study explored how 16 first-year undergraduates describe and think about their multiple social identities. The software influenced both the methodologies and analysis schemes used by the researchers. In these studies, the ability to give a concrete structure to large data sets, code at multiple levels, and pursue iterative analysis methods had many repercussions for the research effort. Appendixes contain the coding structure for two of the studies. (Contains 1 table, 14 figures, and 16 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 431 015

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Jon Margerum-Leys

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Analytical and Methodological Issues in the use of Qualitative Data Analysis Software: A Description of Three Studies

Jon Margerum-Leys (jmargeru@umich.edu)

Jeff Kupperman (jkupp@umich.edu)

Kristen Boyle-Heimann (kpboyle@umich.edu)

The University of Michigan

A paper presented at the American Educational Research Association annual meeting in Montréal, Canada, April 22, 1999.

This research is supported in part by a Spencer Foundation Research Training Grant to the University of Michigan.

TM029847

BEST COPY AVAILABLE

Abstract

The use of data analysis software in the process of qualitative research has become commonplace. Use of such software has clear technological implications in terms of knowledge which must be acquired by researchers in order to use the software. In addition, computer hardware and software have come to be seen as necessary for the conduct of qualitative research. We believe that in addition to technological implications, there are methodological and analytical consequences for the use of qualitative data analysis software. In this paper, we present perspectives gained from the conduct of three qualitative research studies. While the studies differ in their theoretical frame, areas of interest, and scope, their common employment of a particular data analysis software package give us a means by which to explore issues which result from the use of the software.

Computers and specialized software for data analysis have had a far-reaching impact on qualitative research. Richards & Richards (Richards & Richards, 1994) call the potential implications of the use of qualitative data analysis software "dramatic" (p. 445). Bogdan & Biklen (Bogdan & Biklen, 1992) call the use of computers to analyze qualitative data "The greatest change in research technology in the last decade-and-a-half" (p. 181). This change, however, is seen by Bogdan and Biklen as "more technical than conceptual" (p. 27). While technological affordances and impediments may be the most conspicuous issues, we[1] believe that the conceptual issues involved in using software to aid in data analysis are worth consideration. Unpacking these issues may help future qualitative researchers as they make study design choices and decisions regarding the use of computer software as an aid to their analysis.

To explore the conceptual consequences of the use of data analysis software, we present perspectives from three current qualitative studies which have in common their use of a particular data analysis software: QSR's NUDIST. Our interest is not centered on technological concerns; we focus instead on methodological and analytical issues raised by the affordances and constraints of the software used in the data analysis process.

In this paper, we write about "methodological" and "analytical" issues. For our purposes, methodological issues are those which affect the design of a study or which affect how data collection is carried out. Analytical issues are those which impact the analysis phase of a study, after data has been collected; in the methodologies employed in the three studies, analysis and writing phases of the research process were somewhat permeable. More will be written about this later in the paper, but for definitional purposes, analytical issues in this paper may occur during the analysis of data and/or during the writing of study results.

Affordances, as we use the term here, are facets of the analysis software which encourage certain courses of action. This definition grows out of the ecological psychology use of the term (see (Gaver, 1991), but in this paper we consider the consequences of affordances in addition to the affordances themselves.

Data representation, like software affordances, has methodological effects (Larkin & Simon, 1987). The ability of analysis software to represent data in certain forms causes the researcher to structure her methods along lines allowed by the representational capacities available. Additionally, some types of results are more easily represented and thus more easily seen than other types; use of the software may highlight trends which would otherwise be considered inconsequential and may hide trends which otherwise might be quite important. Thus, analysis which is supported by a particular software tool differs both from analysis not supported by software and by analysis supported by other tools.

[1] In sections which refer to issues common to the three studies, "we" refers to this paper's three authors. "I" is used in the sections describing each individual study and refers to the study author.

This paper focuses on the methodological and analytical consequences of the use of NUDIST in three educational research studies. For each study, we discuss the choices which data analysis software allows and encourages, commenting on the methodological and analytical effects of these choices. The variety of conceptual frames and data types represented by the three studies combined with the common use of the same analysis software allows ample opportunity for comparison and contrast among the studies. Data structure, coding hierarchy, flexibility in coding, and reporting capacities are examined in each study, yielding a depiction of the methodological and analytical influences of this particular qualitative data analysis software tool.

QSR NUDIST Described

Before describing the studies and considering their respective methodological and analytical themes, this section briefly describes QSR NUDIST, the software we used for data management and analysis. This brief overview is intended to familiarize readers with the basic capabilities and modes of operation of the software.

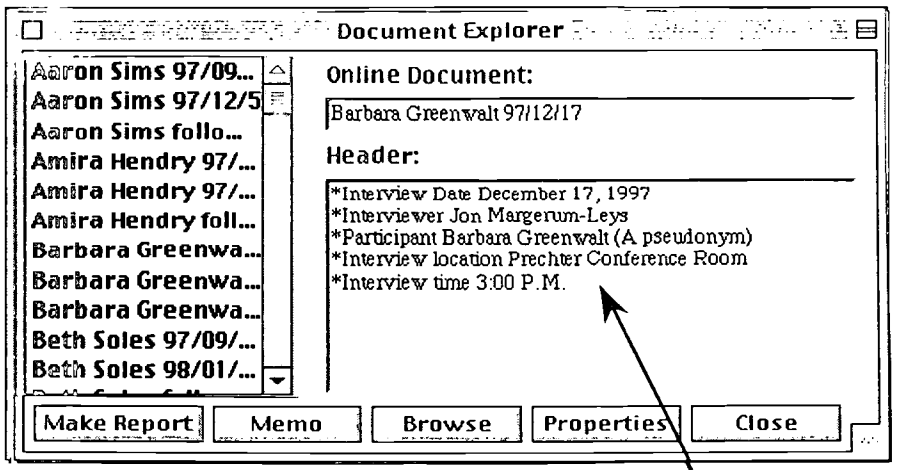
At the broadest level of organization, all of the data in a particular study are contained in a "Project". Within the project are the raw documents which comprise the data, the coding structure, and assorted notes and memos created by the researcher in the process of managing the data. In the subsections below, we illustrate the organizational components and metaphors used by NUDIST in constructing its projects.

Document Handling

NUDIST's ability to allow a researcher to manage large numbers of documents is one of its prime attractions. Figure 1 below shows the simple system used to display the documents within the NUDIST project. Over three hundred pages of transcripts are represented by the list of documents at the left side of the window in the illustration below. NUDIST combines the accessibility to text of a word processing program with the data handling capacity of a database program, as well as some of the reporting features of a spreadsheet.

Figure 1: Document Explorer Window

The seventeen documents which comprise the data for study 1 are listed here.



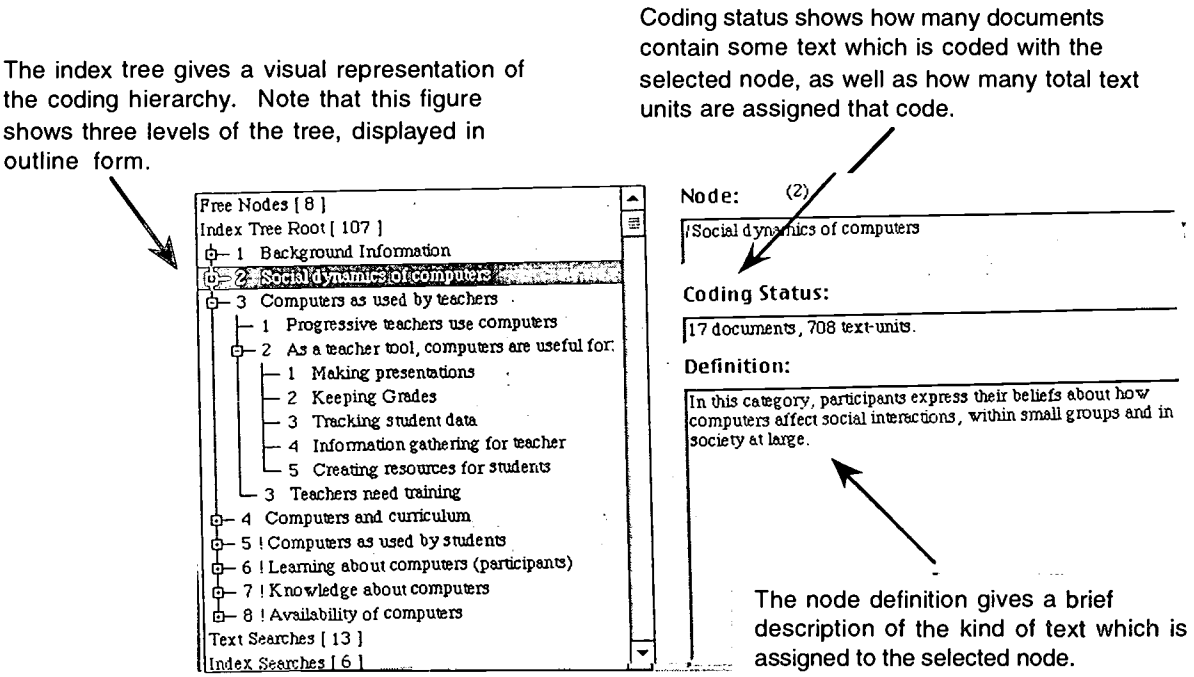
The researcher determines the kind of header information for each document in the system. In this example, the date, time, place, and participants are listed for the selected interview.

The Node System

Text which is entered into NUDIST can be organized and analyzed through the creation and use of a coding system; this coding is one of NUDIST's primary functions. The coding system, in turn, is structured through a system of nodes. Each code, whether it is thematic (relating to the content of the data) or structural (relating to the structure of the data) is assigned to a node on the coding tree. These nodes can be hierarchically arranged, as shown in Figure 2 below. Alternatively, nodes can be simply listed in a non-hierarchical structure such that all nodes are at one level. In addition to showing the structure of the nodes, the node explorer shown in Figure 2 gives some basic information about the definition of the selected node ("social dynamics of computers" in this case) and the amount of text in the project which has been assigned to that node.

BEST COPY AVAILABLE

Figure 2: Node Tree



Working With Individual Documents

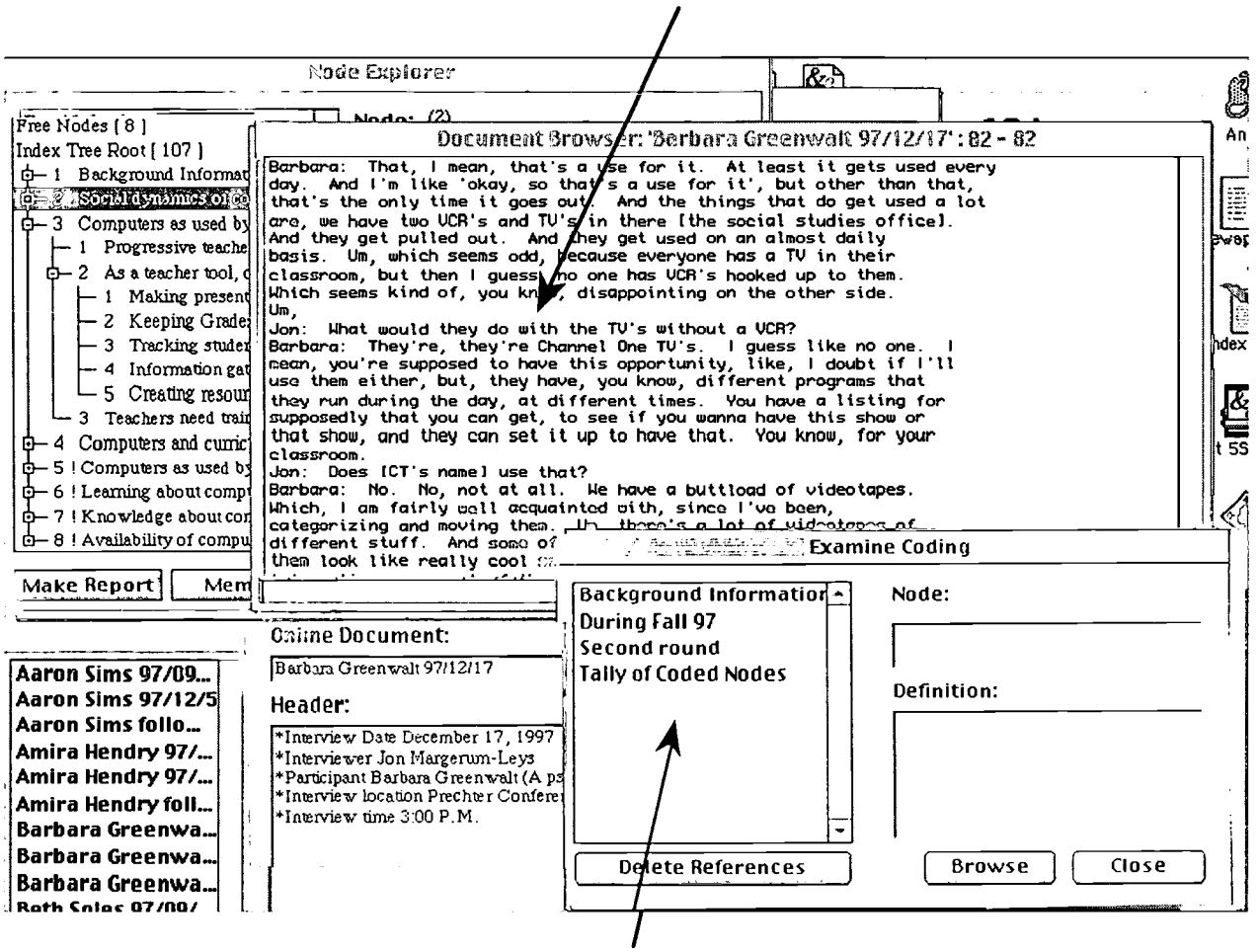
Figure 3 below shows an individual document from study 1 of this paper in the process of being browsed. In the figure, the text has been coded, though the coding is not immediately visible. The full text of the interview can be seen and the coding of any particular piece of text is accessible in a separate small window. In the coding process, any piece of text can be assigned to one or more codes, each of which is a node in the coding tree.

Note that the selected line of text in the figure below is assigned four codes: 'Background information,' 'During Fall 97,' 'Second Round,' and 'Tally of Coded Nodes.' The first two codes are thematic, dealing with the content of the text. Text coded as 'Background information' concerns statements by a participant concerning her background; 'During Fall 97' is the code for statements about a participants' experiences during that particular period of time. The second two codes are not thematic; rather, they are used structurally for data management. 'Second round' is a code assigned to all data from the second round of interviews. By assigning this code, the researcher could quickly separate out data from this round; combining thematic and management codes allows the researcher to find, for example, all statements made in round two referring to background information. 'Tally of Coded Nodes,' the final code shown below, was a very general code assigned to all text which had any coding assigned to it. This allowed the researcher to quickly calculate the percentage of data captured by the coding system.

BEST COPY AVAILABLE

Figure 3: Browsing and Coding

The full text of an interview from study 1 is shown. In this example, each line of text forms a unit which can be coded separately.



The particular codes assigned to the selected line above are shown here.

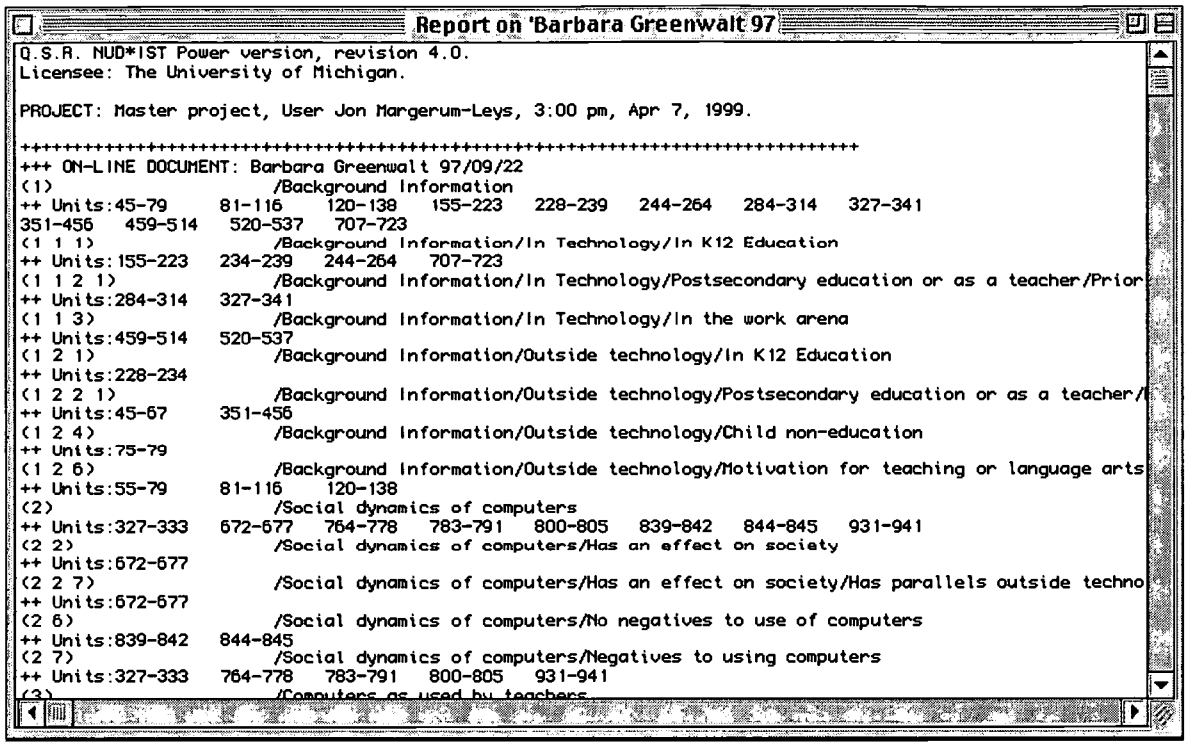
Reporting Coding Results

A variety of flexible reports can be generated from NUDIST at any point in the coding process. From a methodological standpoint, this encourages an iterative approach to the coding process; since reports can be quickly generated, it is easy for a researcher to check on the nature and effectiveness of her coding structure repeatedly during the coding process.

Reports can be created from two main perspectives: Code centered and document centered. Document centered reports, as shown in Figure 4, allow an overview of the thematic contents of a particular piece of data. Code centered reports, shown in Figure 5, allow views of a theme across multiple pieces of data.

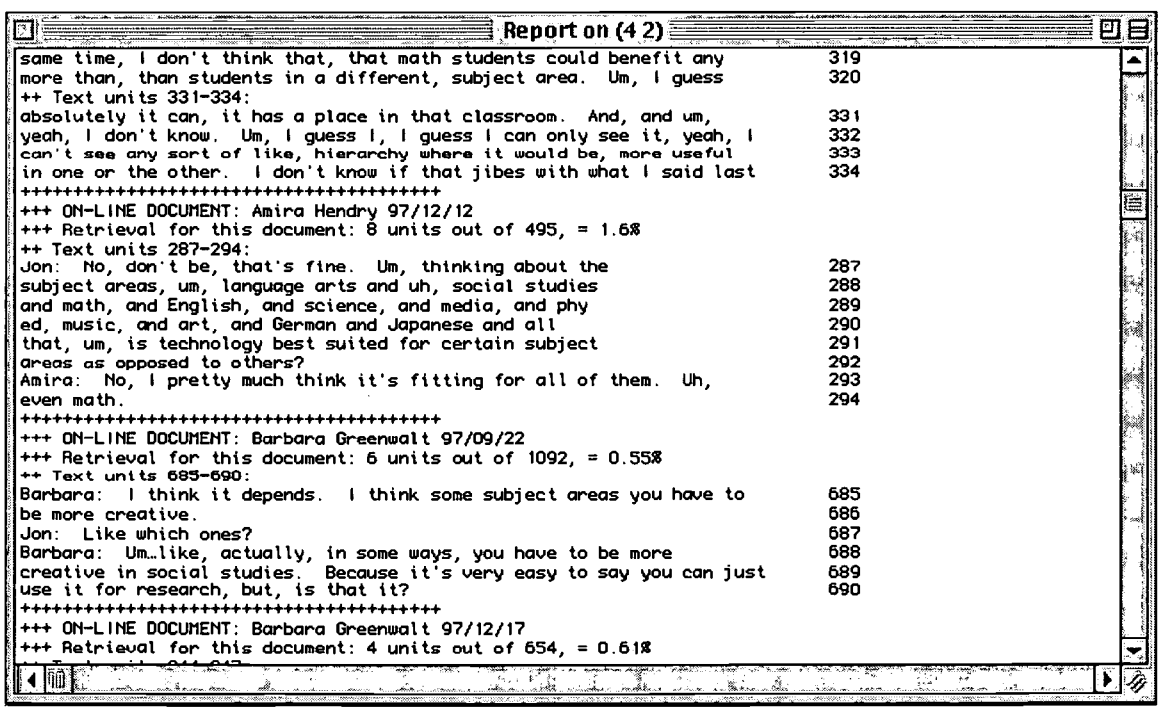
BEST COPY AVAILABLE

Figure 4: Document Report



In the illustration above, the codes assigned to a particular document are shown. By referring to the node tree, the researcher can see which themes can be found in this interview.

Figure 5: Code Report



BEST COPY AVAILABLE

The report illustrated above shows the percent of text coded for the particular code being reported. For example, the interview "Amira Hendry[2] 97/12/12" has 1.6% of the interview assigned to this code. Following the percentage is the full text coded, along with the line numbers at which the text appears.

Ease of Use Concerns

NUDIST offers a powerful, flexible system for qualitative data analysis. Its capabilities facilitate an iterative, bottom-up approach to coding textual data. However, the learning curve for this software is quite steep. Even when the program has been learned thoroughly, its counterintuitive modes of operation can be frustrating. Two of this paper's authors (Margerum-Leys and Kupperman) hold Master's degrees in educational technology; both consider themselves reasonably skilled technologists. Kupperman has taken part in a day-long seminar on NUDIST. Still, using NUDIST remains a challenge for all three of this paper's authors.

Is NUDIST's difficulty-of-use an analytical concern? Perhaps. It might be argued that ineffective use of NUDIST could lead to problematic data analysis. Incorrect coding of a data set or misreading of a NUDIST report could lead to erroneous research conclusions. For example, consider Figure 4 above. It is not easy to interpret the results of the report shown; the format of the report is not conducive to easy interpretation, and no graphic representation of the data is available. It is conceivable that a researcher who had difficulty generating the appropriate report, or who was unable to interpret the reports created by NUDIST, could reach conclusions not warranted by her data. This would be true of any researcher using any analysis system; however, NUDIST's difficulty of use highlights this potential problem.

Conclusion of This Section

Though we have reservations about the difficulty of learning and using NUDIST and believe that these difficulties may have analytical antecedents, NUDIST offers affordances which justify its use. Its ability to manage large amounts of data while retaining access to the data itself make it a useful tool. The sections above give the reader a sense of the capabilities and modes of operation offered by NUDIST. We use this description as a means to help the reader visualize the relevance of the methodological and analytical issues explored in the sections below.

Study Descriptions and Methodological and Analytical Issues Addressed

In the sections below, we describe and discuss the three studies which contribute to this paper. For each, we describe the research setting, methodology, and analysis used. These descriptions serve to create a picture of the studies as a basis for describing the analytical issues entailed in using NUDIST for data analysis.

[2] A pseudonym

Study 1: Teacher Education Students' Beliefs About Technology

Study author: Jon Margerum-Leys

Study Description

This interview-based study examined the beliefs of nine teacher education students at a major research university. The participants were members of a year-long field-based teacher preparation cohort, in which teacher education students earn a Master's degree in education and a teaching certificate. A complete description of this study's participants and methods, along with the study results, appeared as an AERA poster session (Margerum-Leys & Marx, 1999).

The interviews which comprised the study data were loosely structured around protocols consisting of approximately one dozen questions, with follow-up probes for each question. While the same protocol was followed for each interview, conversations varied somewhat due to the individual interests and experiences of the participants.

Interview times ranged between 35 minutes and 75 minutes. Audio tapes were transcribed word for word, including non-word utterances. Transcribed interviews were coded using verbal analysis coding, an analysis method suggested by Chi (1997).

Verbal analysis coding is a method by which a researcher's subjective impression of the data is formalized, combined with an analysis which examines the frequency with which items identified as fitting the formal pattern occur. Creation of the coding and coding categories is "bottom-up," arising from the data rather than a theoretical construct.

The verbal analysis process yielded eight coding categories: Background information, social dynamics of computer-based technology, computers and curriculum, computers as used by teachers, computers as used by students, learning about computer-based technology, knowledge of computer-based technology, and availability of technology. The categories in turn were parsed into 84 distinct codes. In this study, the complex coding tree represents my subjective picture of the data. NUDIST was used to map the 13,000 lines of transcribed material onto the coding tree. Using this structure, the data were examined for patterns within and across codes and categories, participants and interview sets.

Methodological and Analytical Issues

For this study, NUDIST supported an iterative, bottom-up approach to data analysis. While it is possible to accomplish bottom-up coding without the use of data analysis software, the ability to create complex coding structures which could be easily modified encouraged me to pursue coding in which the structure arose from the data. Several issues follow from this:

Consideration of analysis capabilities during study design

When designing the study, I considered among surveys, structured interviews, and semi-structured interviews as sources of data. Surveys and structured interviews offered a built-in starting place for a coding structure; a code could be established for each survey item or interview question. Comparison between participants would be straightforward, with participants' answers matched on an item-by-item basis. However, I felt that a semi-structured interview would give participants a greater chance to express themselves. In turn, if the questions I asked were less structured and avoided leading the participants in too-specific directions, I would have more basis for a claim that interview responses reflected participants' beliefs.

NUDIST's ability to flexibly facilitate thematic and structural coding was an influence in my selection of semi-structured interviews as a data source. Creation of a data structure was not dependant on preset items. The ability to handle a large set of documents, searching and reporting quickly across the set, persuaded me that it would be feasible to draw structure from an ill-structured set of interviews.

Calling attention to data trends

As I generated codes and coded the interview data, I was able to repeatedly explore how often particular codes occurred within the data. In some cases, this led to alteration or rejection of a particular code. For example, codes which occurred infrequently or which occurred only within a particular interview were examined for combination with similar codes or for rejection from the coding structure.

There were instances in which codes arising from the data came to my attention unexpectedly. In one case, I discovered that all participants had mentioned access to technology as a critical element in implementation. None of the interview questions had queried participants on this area and I had not thought it would occur in conversation. However, exploring the developing codes brought to my attention that the topic occurred in every interview. Without the ability to quickly and flexibly examine my developing data structure, this data trend might have gone unnoticed.

With eighty-four codes in eight distinct categories, tracking the relative counts and percentages of each code and category was a formidable challenge. The capacity to do so allowed me to carefully look for trends in the data. While the data in this study were not quantitative, tracking the counts and percentages of each code gave me a numeric sense of a very complex data structure. Switching back and forth between the quantitative nature of counts and percentages and the qualitative data itself allowed a rich combination of perspectives on the data.

Counts and percentages as a means to justify reporting trends

Verbal analysis coding (Chi, 1997), the analysis system used in this study, brings some quantitative justification to qualitative data. Codes which occur across several participants, recur

from interview to interview with a particular participant, or which account for a significant percentage of the data, can be justifiably deemed worthy of reporting.

Consider the example above, in which participants talked about access to technology as a key element of their beliefs about how technology is and should be used in education. By using NUDIST, I was able to confirm that all of the participants had alluded to this issue, and that for several, it was a major focus of their beliefs about technology. Access to technology as an issue accounted for an average of 4.8% (range 0.38%-13%) of the interviews. The ability to have some justification for inclusion of a category such as this, combined with the ability to closely examine incidences of participants talking on a particular subject, gave me confidence in my analysis of the data.

Making the data structure explicit

Within the NUDIST software, it is possible to see a visual display of the developing coding tree. In concert with Inspiration software's Inspiration, NUDIST also supports the creation of more easily-grasped concept maps of the coding structure. Appendix A contains the concept maps for the eight data categories in this study. Creating rough drafts of these maps to represent early iterations of the coding structure was the work of minutes. For me, these early maps represented an epiphany in the analysis portion of this study. Like the blind men and the elephant (Saxe, 1968), I had seen portions of the data but had never been able to step back and view the big picture. Multiple perspectives on data are invaluable in the analysis process. With the visual array provided by the maps of the coding structure, I had three perspectives: The mass of qualitative data itself, in which the participants' voices could be heard; the quantitative look provided by the counts and percentages reporting; and an impression of the skeleton of the body of data presented by the maps. This last was invaluable in allowing me to see the shape of the data as it developed.

Potential criticism of over-rating the validity of the structure

NUDIST allowed me to create a complex analysis structure which is nonetheless comprehensible. This structure seems to me to accurately represent the data. As I see it, the coding arose directly from the data and the structure exists as it does as a result of the nature of the underlying data. It is conceivable, though, that I have created a castle made of straw. Straw, left to its own devices, would never resemble a castle. Perhaps the data in the study under consideration do not really coalesce into the structure I have created.

When using NUDIST, it is easy to be lulled into believing that one's own initial coding is correct. Once coded, data can be mined for an array of very convincing reports and maps. These appear to validate the researcher's analysis. At the level of counts and percentages and visual maps, though, what is being reported is not the data itself, but the subjective structure created by the researcher's own coding system. Without repeated, careful looks at the data itself, the analysis is buried under multiple layers of structure which only have as much validity as the correspondence between the codes and the underlying data.

My response to this potential criticism comes on two levels: Ethical and technological affordance. Ethically, any researcher following any analysis scheme of any data set has an obligation to do so thoroughly. I would concede that NUDIST makes it easier for a researcher to fool herself. Still, the ethical obligation mitigates effectively against the creation of fatally flawed data analysis structures.

From a technological affordance view, NUDIST makes it very simple to include the data itself into any coding report. In the example given above, when generating a report on how often participants expressed concern about access to technology, I could include the text of statements given that code, allowing me to re-examine the appropriateness of the coding.

Study 2: Academic, Social, and Personal Uses of the Internet: Cases of Students From an Urban Latino classroom

Study author: Jeff Kupperman

Study Description

This study is based on data from a project where the families of middle school students in an ESL (English as a Second Language) science class were given home Internet technology as part of a project-based unit on the respiratory system (Fishman, Kupperman & Soloway, 1999). The technology, dubbed "NetTV" in the study, was a relatively inexpensive device that when connected to a television set and a phone line allowed one to browse the World Wide Web and use e-mail. The families, who were Mexican and Puerto Rican immigrants living in an inner-city neighborhood, had never had home access to the Internet before. Taking a sociocultural view of technology adoption (Bruce & Hogan, 1998; Kaptelinin, 1996; MacKenzie, 1996; Nardi, 1996; Schement, 1995), the study focused on the ways that children and their families perceived the technology and used it for academic, social, and personal purposes. At the center of the study are case studies of four students in three families (two students were siblings).

Data sources

Data included structured interviews with 14 students in the third month of the project and follow-up interviews with three target families at the end of a year; fieldnotes from the classroom and home visits made by project staff throughout the project; electronic logs of Internet use by each family; and notes taken from 4 selected days of classroom video. A total of 47 text documents, many containing data from multiple days, were imported into the NUDIST project.

These data were used to construct cases studies that explored use of the technology in three categories: "family life," which included interactions among family members around NetTV; "school life," which included use of NetTV for schoolwork, and "kid life," which included the ways NetTV was used for social and expressive activities. Home use of the Internet by first time, "non-mainstream" users is an area where little research has been done (Kupperman, 1999), and the case studies aimed to generate broad themes, rather than to test particular theories.

Methodological and Analytical Issues

NUDIST was chosen because of the need to organize and process large amounts of textual data, as well as look for patterns across data from different sources. NUDIST supported an iterative processes of data display and analysis (Huberman & Miles, 1994) that was partly top-down (derived from a theoretical perspective) and partly bottom-up (generated from within the data): The research design and research questions led to some initial categories, but categories were also generated and refined during repeated passes through the data. For example, interviews were structured enough that the responses could be pre-processed in a form that allowed NUDIST to automatically create a category for each question. Each target student also had an a priori category to tag all data pertinent to that student. Other codes were created "on the fly" during read-throughs of the data or text searches; these were often combined into larger categories later. The three large categories of family life, school life, and kid life were created relatively early, and later coding was done with these in mind.

Because the study was exploratory and broad in scope, it was sufficient to identify an area of text that corresponded to a particular theme, as opposed to comparing or counting linguistic features of the text. Therefore, no special attention was paid to the size or structure of text units; text units were defined as the default of one word processor line.

Also, no special effort was made to make sure the final codes were strictly consistent and hierarchical. For this paper, a choice was made to do initial organization of data in NUDIST, but to do further organization and synthesis through working directly with narrative drafts of the case studies. During the writing of the case studies, the NUDIST data file was used to check evidence and provide examples, and some new coding was done at this time. However, as the writing progressed, use of NUDIST became focused less on manipulating the index system and more for retrieving evidence to support arguments in the writing. This choice was partly due to personal work style of the author, but it is also due to the fact that much of the data was useful for suggesting themes rather than making rigorous comparisons. For example, there was ample evidence that one focus student had a particular interest in a website for Lowrider Magazine and viewed it again and again. In another example, classroom video captured a conversation between several students about looking at the Playboy web site, which was suggestive of the way some of the students viewed the Internet. In this case the students' conversation highlighted a difference between one student who saw the Internet as a way to challenge authority and do grown-up things, and another student who saw it more as part of her schoolwork. However, since the data did not capture every single conversation between students, it was impossible to say, for example, how common these conversations were. It was enough to have this piece of data in the NUDIST database, and less important how it fit into the coding tree.

Impact of multiple-level coding

An important feature of NUDIST for this study was the ability to code any particular segment of text in multiple ways, then call up text segments that represented an intersection of certain

codes. In particular, it was possible to code separately by family and by theme, and compare themes for each family. For example, all text that pertained to the student Jorge was coded "Jorge," and all text that had to do with the use of e-mail by parents was coded "parent e-mail." To examine how Jorge's parents used e-mail, it was possible to search for the intersection of "Jorge" and "parent e-mail." This flexibility made it easy to add, merge, and refine codes, and then quickly examine those codes for each case. As mentioned above, the NUDIST database was used to retrieve evidence and check hypotheses during the writing of the case studies, and so it was important to be able to do these kinds of searches quickly and easily.

Study 3: Multiplicative Identities: First Year Undergraduates' Explorations of Their Own Multiple Identities

Study author: Kristen Boyle-Heimann

Study Description

The purpose of this study is to explore how first year undergraduates describe and think about their multiple social identities. The term social identity, drawn from the social justice literature, refers to identifying as a member of a particular group based on one's race, religion, ethnicity, gender, sexual orientation, socio-economic class, ability or age (Boyle-Heimann, 1997). I collected data over the course of one academic year acting as an observer in a course in which all of the participants were enrolled. Two interviews were conducted with each participant in the first term and two in the second term, culminating in a final interview and questionnaire at the end of the academic year. Primary data include four audio-taped transcribed interviews with each of the sixteen study participants, responses to a series of three written questionnaires, and participants' written work. In addition, individual and group videos and fieldnotes from observations of class sessions provide further information and supporting evidence. Multiple data sources provide depth and allow for potential triangulation of emerging themes.

NUDIST was selected as an analytic tool after the study had been designed and during the process of data collection. Thus, the choice to use NUDIST did not impact the study design and has no methodological implications for this study.

As I undertook both issue-focused and case-focused analysis (Weiss, 1994) to examine themes across cases as well as to explore the richness and depth of particular cases, I employed NUDIST to code and examine data. Once documents are coded and the codes entered into NUDIST, the software becomes useful in exploring coding patterns and illustrating themes emerging from the data. NUDIST makes it feasible and efficient to do thematic analysis across large quantities of data from non-contiguous sources. Thus I was able to expediently consider a variety of ideas and themes. The use of data analysis software enhances the scope and efficiency of my analysis.

Methodological and Analytical Issues

As described previously, this is an interpretative study exploring how first year undergraduates describe and think about their multiple social identities. Issue-focused and case-focused analysis (Weiss, 1994) examine themes across multiple data sources. The study centers around case studies of three particular students and thematic, issue-focused analysis across data from all of the 16 first year undergraduate students enrolled in the same course.

NUDIST software was chosen for this study due to the need to process and analyze large amounts of data from varied sources (interview transcripts, fieldnotes, participants' written work, and questionnaire responses) and across study participants. I had completed a previous project analyzing field notes, interview transcripts, and questionnaire responses using a word processing program as a slightly updated mode of cutting and pasting all material into coded categories. The piles of paper strewn across my floor were unwieldy in a relatively small study, and I knew there had to be a more effective way of managing the large quantity of data collected for the current study which includes 64 audio-taped interviews and more than 1,000 pages of transcripts.

Effect of text unit size decisions

When initially processing text documents, NUDIST requires the researcher to select segments of data to be coded as text units. A unit of text is recognized by the software each time there is a hard return within a document, and a text unit is the smallest segment that can be coded in NUDIST. While it is possible to combine paragraphs by using line breaks instead of hard returns, it is more common to either code at the paragraph level (with a hard return at the end of each paragraph) or the line level (with a hard return at the end of each line). Any coding scheme requires the researcher to make decisions about parameters, and NUDIST requires this choice be made when documents are set up to be imported into the software so that the hard returns are properly placed to indicate text units.

I did not want to make arbitrary or inconsistent decisions about where a particular unit of data began and ended and did not desire the level of detail associated with coding each word or line of data. Thus, for consistency I decided coding units would be speech units within each interview. That is, a coded unit began each time a speaker started to speak and ended each time the same speaker stopped speaking. For the participants' written work, paragraphs were used as coding units to honor where the participant had decided to begin and end a particular set of ideas. Coding in NUDIST then had to be conducted with these as parameters for text units.

Coding at the paragraph and speech unit was appropriate for the analysis needs of this study. Speech units generally followed a common theme. In rare cases participants switch topics dramatically within a speech unit. In those cases, multiply coding a paragraph allowed the text to be examined from both topics. As I knew I was not doing linguistic analysis or the kind of quantitative/qualitative analysis done in study one, coding at the single line level was not

necessary. Coding at the paragraph level honored participants' sense of where the speech broke while giving a fine enough grain to the text for the analysis purposes.

Iterative nature of coding process

The coding scheme for this study began with an initial line-by-line reading of the entire corpus of data on 16 first year undergraduates collected over the course of one academic year (64 interview transcripts, 48 sets of questionnaire responses, students' papers from first term, fieldnotes from throughout the academic year, and notes summarizing participants' individual and group videos). This initial reading of the data identified themes, patterns, and variations in the data and led to a preliminary set of approximately 550 inductively generated codes (Emerson, Fretz, and Shaw, 1995). The initial set of codes was examined for duplications, common ideas that could be linked under one code, and codes that were too specific or narrow to be broadly useful. This iterative process led to a refined list of approximately 370 codes and subcodes. The refined set of codes was employed to code pilot data from two study participants not selected as case studies. After the pilot coding, I again refined the codes, eliminating duplication and codes that were too specific or narrow (e.g., codes generated by or used only in one small chunk of data), leading to approximately 300 codes and subcodes. As the coding scheme developed, NUDIST provided the flexibility to refine codes by deleting, renaming, and combining codes. These codes were then used to code data from participants selected as case studies. During the process of developing and refining codes, I kept lists of the codes in a Microsoft Word file, using Word to alphabetize the codes, with subcodes indexed under the main code. In addition, during this process I also wrote memos tracing my thinking and rationale (Emerson, Fretz, and Shaw).

Initial coding identified many more themes than could be pursued in one project. Thus, I then focused analysis for each case study on recurrent issues linked to what had emerged as a focal interest in the study (Weiss, 1994) — how participants talk about and describe their struggles around social identities. Codes and coded data had been entered in NUDIST, and now the software facilitated and expedited the process of focused coding and analysis by allowing me to produce and print reports displaying data coded around the focal themes identified. I read and analyzed the NUDIST-generated reports to distinguish subthemes, subtopics, differences and variations within the central topics in a particular case study (Emerson, Fretz, and Shaw, 1995; Stake, 1995; Weiss, 1994).

For example, focused coding of the case of the same male participant previously described might lead to pursuing issues around his description of gender issues in his experiences. Analysis of a report on all data for this case coded "gender" might highlight the theme of this participant's relationships with women as a potentially fruitful area for further analysis. This theme would then be elaborated and explored as the case study is written.

Next, an initial set of guiding questions focused coding of data from the three case studies around issues such as how participants talk about their own multiple identities. This initial focus was broad enough to permit exploration of a wide variation of themes accented in each case. Therefore, it was necessary to employ a large number of codes and subcodes. NUDIST

facilitated the use of a large number of codes because the software made it possible to effectively manage and display data coded with numerous codes. The flexibility to manage and utilize a large number of codes is thus a critical affordance of NUDIST allowing the researcher to identify and explore a wide variety of themes and patterns.

I used NUDIST to code, sort, select, and display data. Coded data was then employed in thematic analysis both within and across cases (Weiss, 1994). During the processes of developing initial guiding questions for the study and then collecting data, I wrote memos to track the evolution of my thinking around the ideas, themes, and issues and recorded my insights on topics and categories identified during data collection as areas of initial interest and areas to pursue further (Emerson, Fretz, and Shaw, 1995). Initial data analysis followed the patterns of open and focused coding suggested by Emerson, Fretz, and Shaw. First, after all of the data had been collected, I read through the entire corpus of data collected line by line to elaborate and expand themes and patterns identified in memos and to identify any new themes not previously recognized. I then employed open coding, reading interview transcripts, participants' written work, fieldnotes, and notes on participant-produced videos, and continued to identify and formulate themes and ideas to both distinguish new and disparate themes and to explore patterns and themes cutting across data.

Analysis affordances and constraints

A particularly useful feature of NUDIST's numerous options in how to sort coded data for analysis is the ability to code an entire document with "base data" and "case data" codes. Base data, referred to as structural codes in study one above, in this study included demographic information about the participants (gender, race, religion, etc.) and the type of data (interview, participants' written work, etc.). This permits examination of a particular category of data. For example, I might wish to examine what all of the Jewish women in the study said about religion by examining data that was coded with base codes "female" and "Jewish" and the specific code "religion." It is quite expedient to explore data from varied data sources (e.g., interview transcripts, participants' papers, and fieldnotes) and from varied participants. It is also possible to explore alternative themes so that analysis is not restricted by time limitations in the same way it might be with more traditional methods of data organization and processing. The exploration of alternative themes allowed me to evaluate which themes identified by coding and analysis were fruitful for further analysis and which were not important or relevant to the particular study at hand.

For example, in one case study the participant ranked sexual orientation as an extremely important identity in the questionnaire responses, yet my sense from coding the interview transcripts and writings for this participant was that sexual orientation rarely came up other than in the questionnaire. To explore to what extent and in what ways the participant talked about sexual orientation in interviews and written papers, I generated a report in NUDIST of all excerpts from this participant's documents coded as "sexual orientation." The report supported my hunch that the participant rarely mentioned sexual orientation. It may be relevant to point to

and comment on this finding in writing up my analysis, but the finding also suggested that there was little data I could use to directly explore the issue of sexual orientation in this case study.

The reporting capacities of NUDIST again provide flexibility. Numerous codes can be combined in one report to explore a particular theme. For example, to explore a male participant's relationships with women, I might look at all of the data coded for this case with the codes "gender," "friends – gender and friends," "romance," and "social life." Just examining the data coded with one of those codes would not capture the scope of data I desire. If I later decide that it would also be useful to look at this participant's relationship with his mother in exploring his relationships with women, it would be easy to simply add the code "family – relationship with parents" and generate a new report. I am not hesitant to explore or try new themes and ideas because it is easy and quick to gather and display the relevant coded data. Ultimately, the flexibility and expediency NUDIST allows may afford a more thorough, comprehensive, and sophisticated analysis than would be feasible with traditional methods.

Efficiency of software use

As with all research methods and tools, the use of NUDIST brings not only benefits but constraints. For example, a colleague pointed to a potential constraint in using NUDIST in that the researcher does not have to go through the process of sorting the data by hand (or by cutting and pasting data in a word processing program) into categories. I have been asked if I am missing something by using software to sort and collate data for me.

The decision about what data to sort and how to sort it is still mine. As noted in the example above, in order to explore the participant's relationships with women I had to make decisions about what type of data and what codes to examine. In addition, I actually found my analysis to be more effective than when I used a more traditional method for data sorting and processing in a previous study because I could spend my time exploring varied themes rather than spending countless hours sorting data.

Technological capabilities and opportunities as an analytical issue

Finally, although the current paper focuses on conceptual rather than technical issues in the use of software to aid in data analysis, it is important to note that the researcher's familiarity and comfort level with technology and software may well impact how extensively she employs the varied modes through which NUDIST can be utilized. The particular affordances and constraints I have discovered using NUDIST may well differ in part from those of the other two authors of this paper due to the differences in our expertise utilizing technology. Kupperman and Margerum-Leys work on research projects and in capacities specifically designed around uses of technology. In addition, they work in a special office area that purposefully provides access to advanced hardware and software. In contrast, I work at home on a relatively simple computer with limited software, not in a technology laboratory. I am not a "techie" as are my two colleagues. Our exposure to and comfort level with software such as NUDIST varies greatly.

The implications of our varied expertise and comfort with technology are that my use of NUDIST is largely confined to using the software in a basic manner to do the type of analysis I would have without the software but in a more expedient and thus more extensive manner, as described above. In contrast to my use of NUDIST to largely make analysis easier and more efficient, the other two study authors are able to use the software to expand the types of analyses they do and can explore the limits of what the software can do to meet their needs.

Conclusion

The three studies which comprise this paper are quite different from each other in area of interest, theoretical frame, methodology, scope, and analysis priorities and methods. These differences allow us to consider multiple perspectives on our common use of NUDIST for data analysis purposes.

Study one draws for its theoretical frame on the literature surrounding teacher beliefs. Methodologically, it is an interview-based beliefs study. As such, its analysis does not make any claims to describing reality; rather, it seeks to describe participants' perspectives and commentary on the role of technology in teaching and learning.

The core of study two is a set of case studies which draw from a variety of data sources related to home use of the Internet by three families. Interviews, field notes, video, and computer log files were used in combination to create the case studies. While each data source by itself provides an incomplete picture of the families' Internet use, together they provide enough evidence to identify themes for each family.

Study three explores first year undergraduates' descriptions and thinking about their own multiple social identities. Methodologically, it is a multi-method interpretive study centering largely around interview transcripts, participants' written coursework, and questionnaire responses. It focuses mainly on social identities but also attends to interlinking social and personal identities acknowledging the ways in which identity is both "personal (residing within the individual) and social (residing within social relationships and society)" (Ashmore and Jussim, 1997, p. vii).

Table 1 below gives an overview of the methodological and analytical issues addressed by each of the study authors.

Table 1: Issues Addressed

Margerum-Leys: Teacher education students' beliefs about technology	• Analysis capabilities as an influence on design • Calling attention to data trends • Justifying the reporting of data trends • Making the data structure explicit • Possibility of over-rating the validity of the structure
Kupperman: Academic, social, and personal uses of the internet	• Impact of multiple-level coding
Boyle-Heimann: Multiplicative identities	• Effect of text unit size decisions • Iterative nature of coding process • Analysis affordances and constraints • Efficiency of software use • Technological capabilities as an analytical issue

The issues addressed above are primarily analytical in nature, as we have defined 'analytical' for this paper's purposes. As NUDIST is an analytical tool, it is not surprising that it is in the analysis phase that we perceive NUDIST to have the most impact. However, these analytical themes in turn impact study design and ultimately have an effect on the reporting and writing phases of research as well. Seen in aggregate, the issues raised in the studies above give a range of examples of the ways in which the use of qualitative data analysis software (in this case, NUDIST) impacts qualitative research in ways that go beyond technological considerations.

Educational Importance

By providing perspectives gained from three studies which represent a breadth of theoretical frames and data types while employing a common data analysis software, we address analytical and methodological issues raised through the use of this software. Affordances of the software influenced both the methodologies and analysis schemes used by the researchers. The ability to give a concrete structure to large data sets, code at multiple levels, and pursue iterative analysis methods have repercussions for multiple phases of the research endeavor.

As researchers consider research design and the role of data analysis software, we hope that the methodological implications of the use of such software are weighed alongside the promise of technical efficiency. While the impact of technology may be more pragmatic than theoretical, the conceptual issues outlined in this paper merit consideration when making methodological and analytical decisions regarding research design, enactment, analysis, and reporting.

Acknowledgements

The authors gratefully acknowledge the input of three anonymous reviewers from the American Educational Research Association Qualitative Research SIG. We also would like to thank our respective advisors Ron Marx, Barry Fishman, and Fred Goodman, professors at the University of Michigan.

References

Ashmore, R. D., & Jussim, L. (Eds.). (1997). Self and identity: Fundamental issues. (Vol. 1). New York: Oxford University Press.

Bogdan, R. C., & Biklen, S. K. (1992). Qualitative research for education. (Second ed.). Boston: Allyn and Bacon.

Boyle-Heimann, K. P. (1997, March). College students' perceptions on interactions across difference. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Bruce, B. C., & Hogan, M. P. (1998). The disappearance of technology: Toward an ecological model of literacy. In M. C. M. David Reinking, Linda D. Labbo, Ronald D. Kieffer (Ed.), Handbook of literacy and technology (pp. 269-281). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. The Journal of the Learning Sciences, 6(3), 271-315.

Fishman, B., Kupperman, J., & Soloway, E. (1999, April). Linking urban Latino families to school using the web: A pilot study. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.

Gaver, W. W. (1991,). Technology affordances. Paper presented at the Conference on Computer and Human Interaction.

Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research . Thousand Oaks, CA: Sage.

Kaptelinin, V. (1996). Activity theory: Implications for human-computer interaction. In B. A. Nardi (Ed.), Context and consciousness: Activity theory and human-computer interaction (pp. 103-116). Cambridge, MA: MIT Press.

Kupperman, J. (1999). Academic, social, and personal uses of the Internet: Cases of students from an urban Latino classroom. Unpublished Unpublished manuscript, University of Michigan, Ann Arbor.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65-99.

MacKenzie, D. (1996). Knowing machines: Essays on technical change. Cambridge, MA: MIT Press.

Margerum-Leys, J., & Marx, R. (1999, April 19). Teacher education students' beliefs about technology. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, CA.

Nardi, B. A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. In B. A. Nardi (Ed.), Context and consciousness: Activity theory and human-computer interaction (pp. 69-102). Cambridge, MA: MIT Press.

Richards, T. J., & Richards, L. (1994). Using computers in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 445-462). Thousand Oaks, CA: Sage Publications.

Saxe, J. G. (1968). The blind men and the elephant. In F. C. Sillar & R. M. Meyler (Eds.), Elephants ancient and modern (pp. 255). New York: Viking Press.

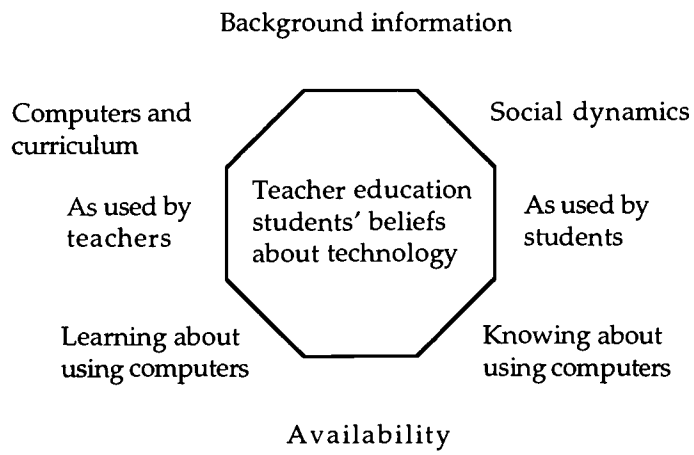
Schement, J. R. (1995). Divergence amid convergence: The evolving information environment of the home, Crossroads on the information highway: Convergence and diversity in communications technologies .

Appendix A: Coding Structure for Study 1

The coding maps below were created by QSR NUD*ST and Inspiration Software's Inspiration. Since the entire map is too large to fit on a single page, I begin with a figure showing the major categories, followed by a separate figure showing each category and its associated codes.

For figures two through nine, the category Figures, the following convention is used: The main topic area for the category is denoted by a darkly shaded rectangle ( ). Subcategories are denoted by an lightly shaded rectangle ( ). Creation of a subcategory occurred when a group of related codes was grouped. For instance, in category 1, "Background Information", separate subcategories were used to group background regarding technology and background outside of technology. Codes themselves are indicated by a shaded oval ( ).

Figure 6: Coding Categories



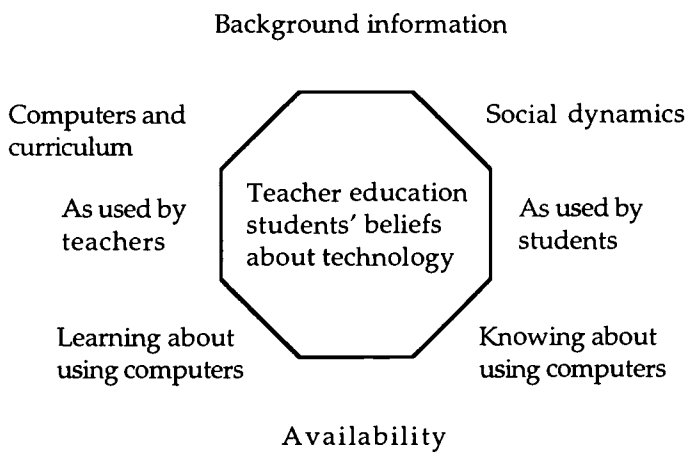
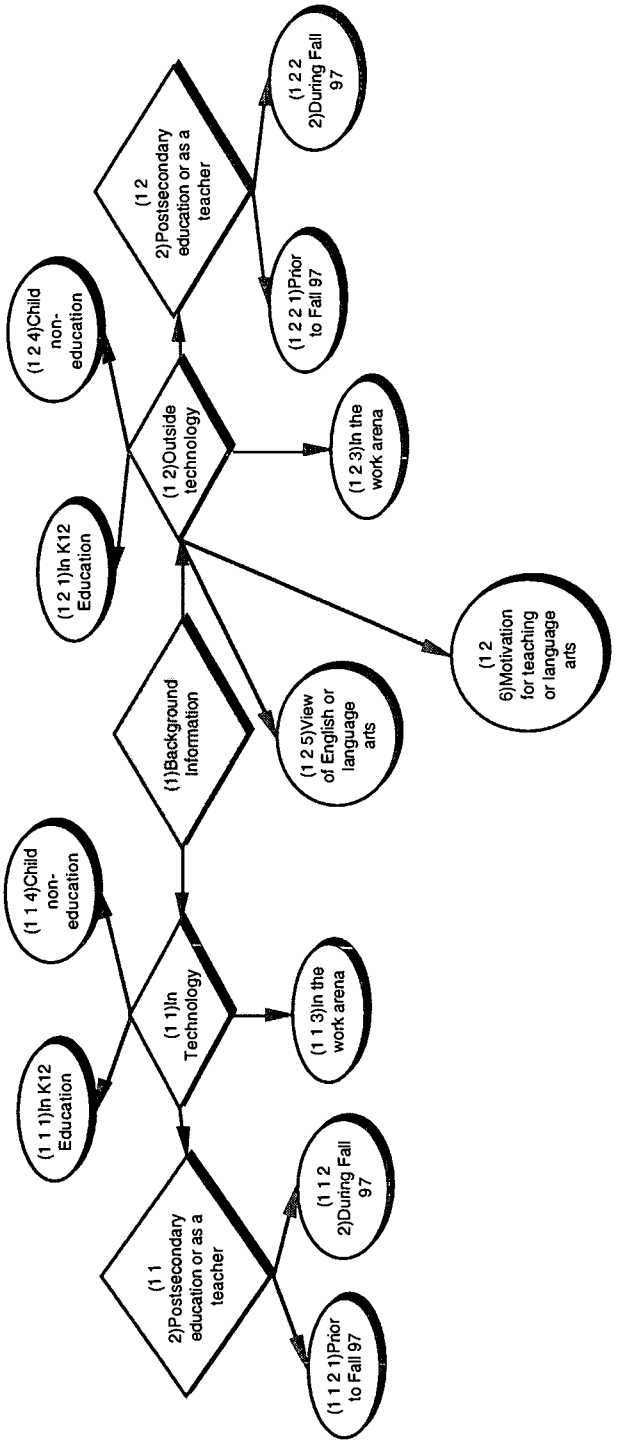
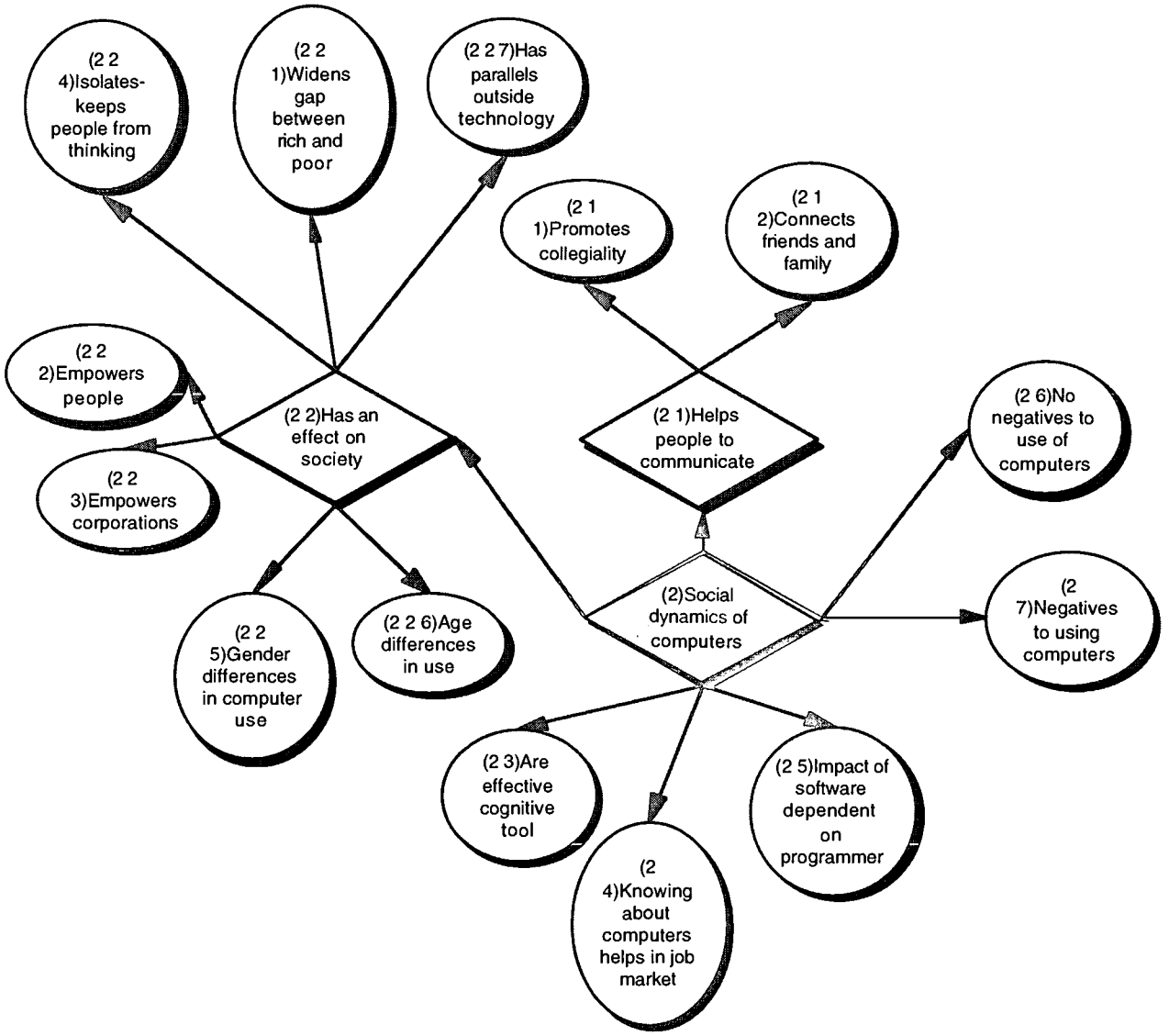


Figure 7: Background Information, 12 Codes



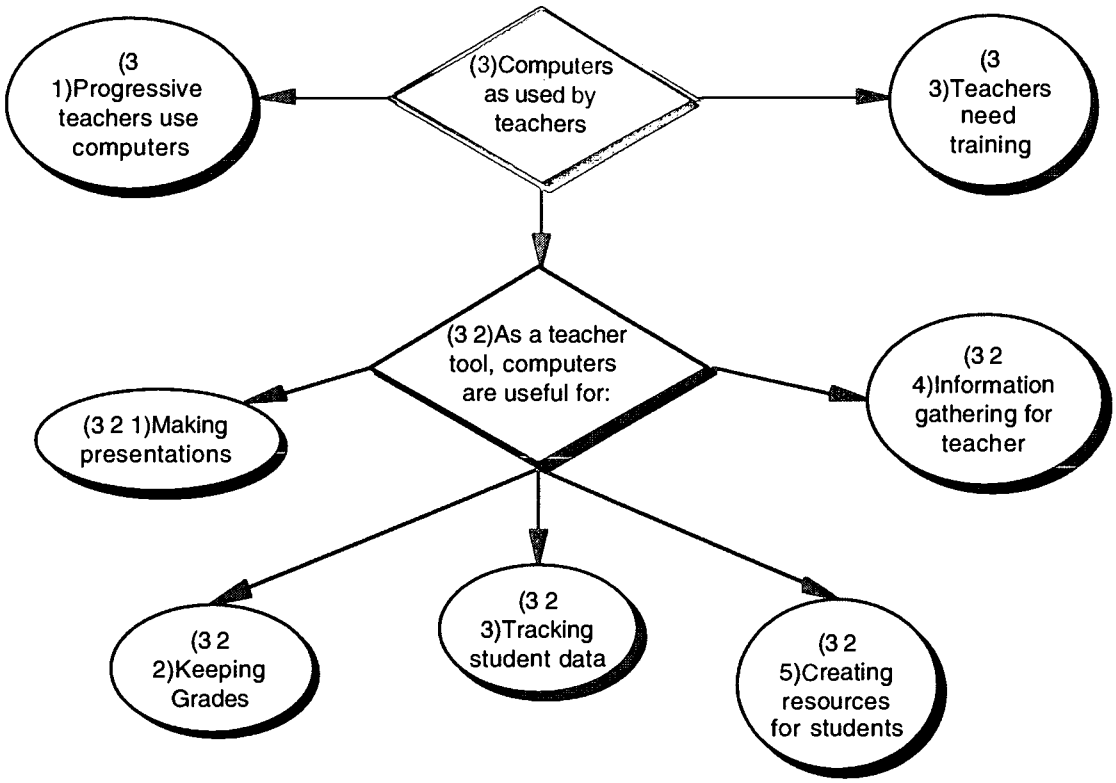
BEST COPY AVAILABLE

Figure 8: Category 2, Social Dynamics of Computers, 14 Codes



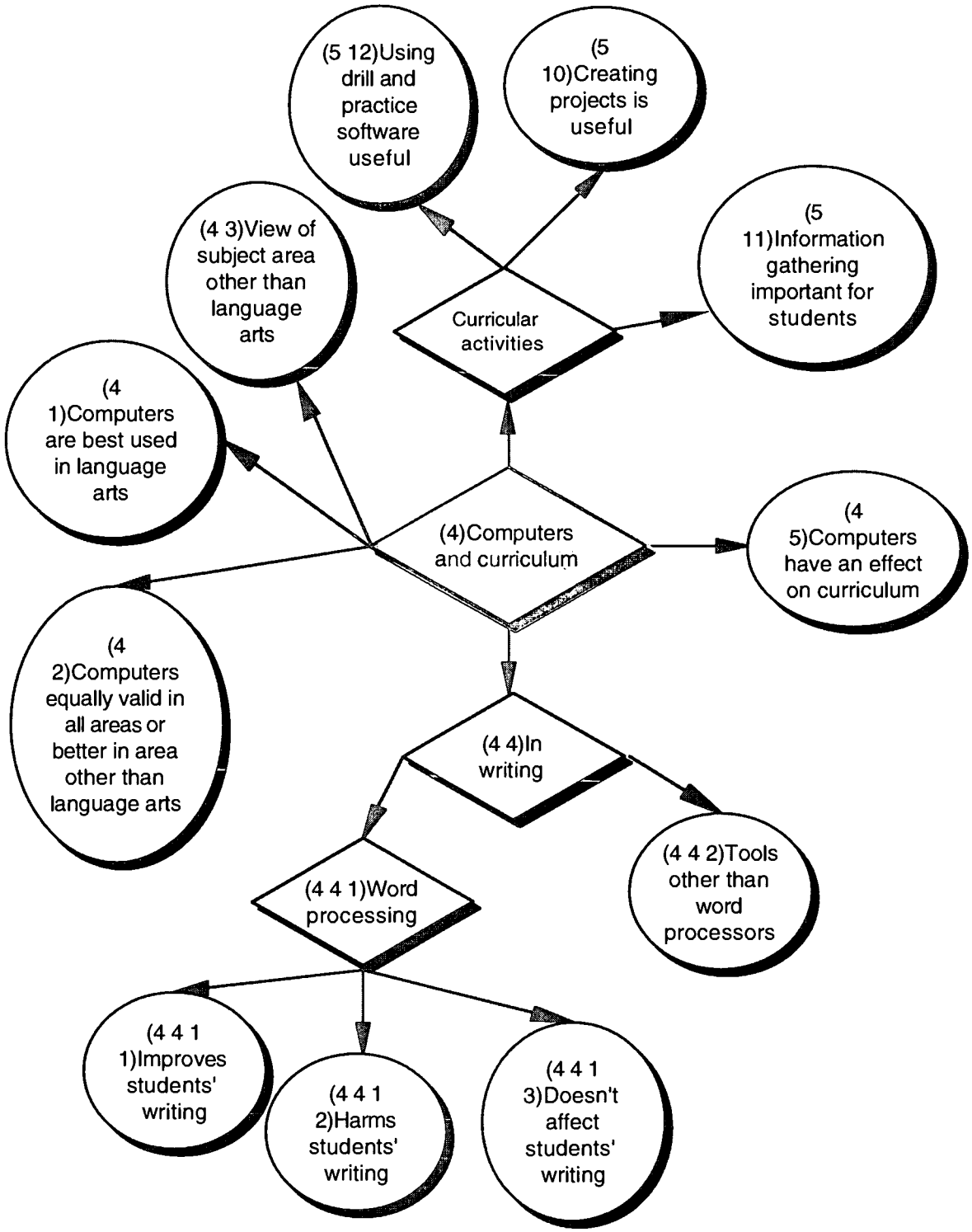
BEST COPY AVAILABLE

Figure 9: Category 3, Computers as Used by Teachers, 7 Codes



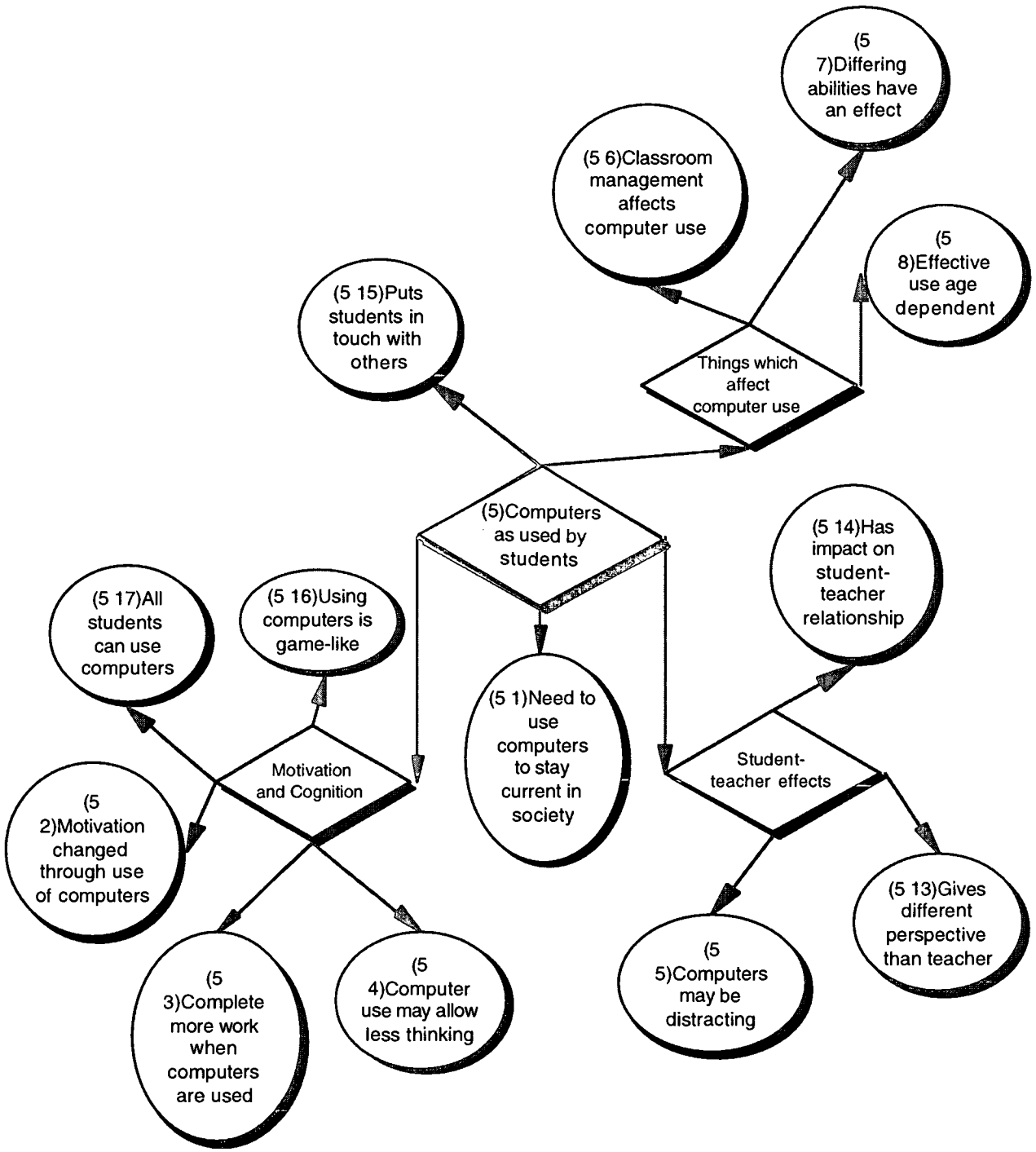
BEST COPY AVAILABLE

Figure 10: Category 4, Computers and Curriculum, 11 Codes



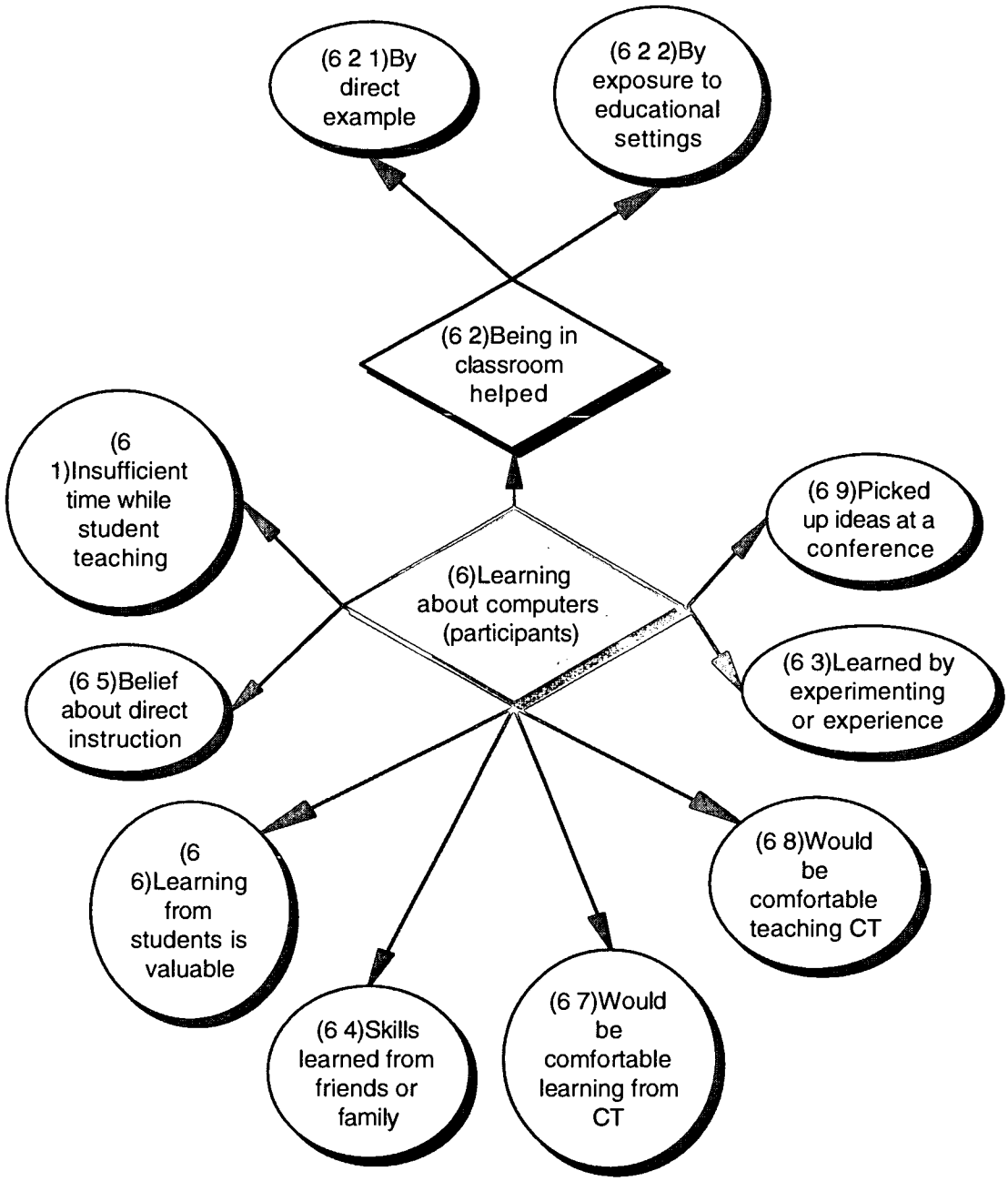
BEST COPY AVAILABLE

Figure 11: Category 5, Computers as Used by Students, 13 Codes



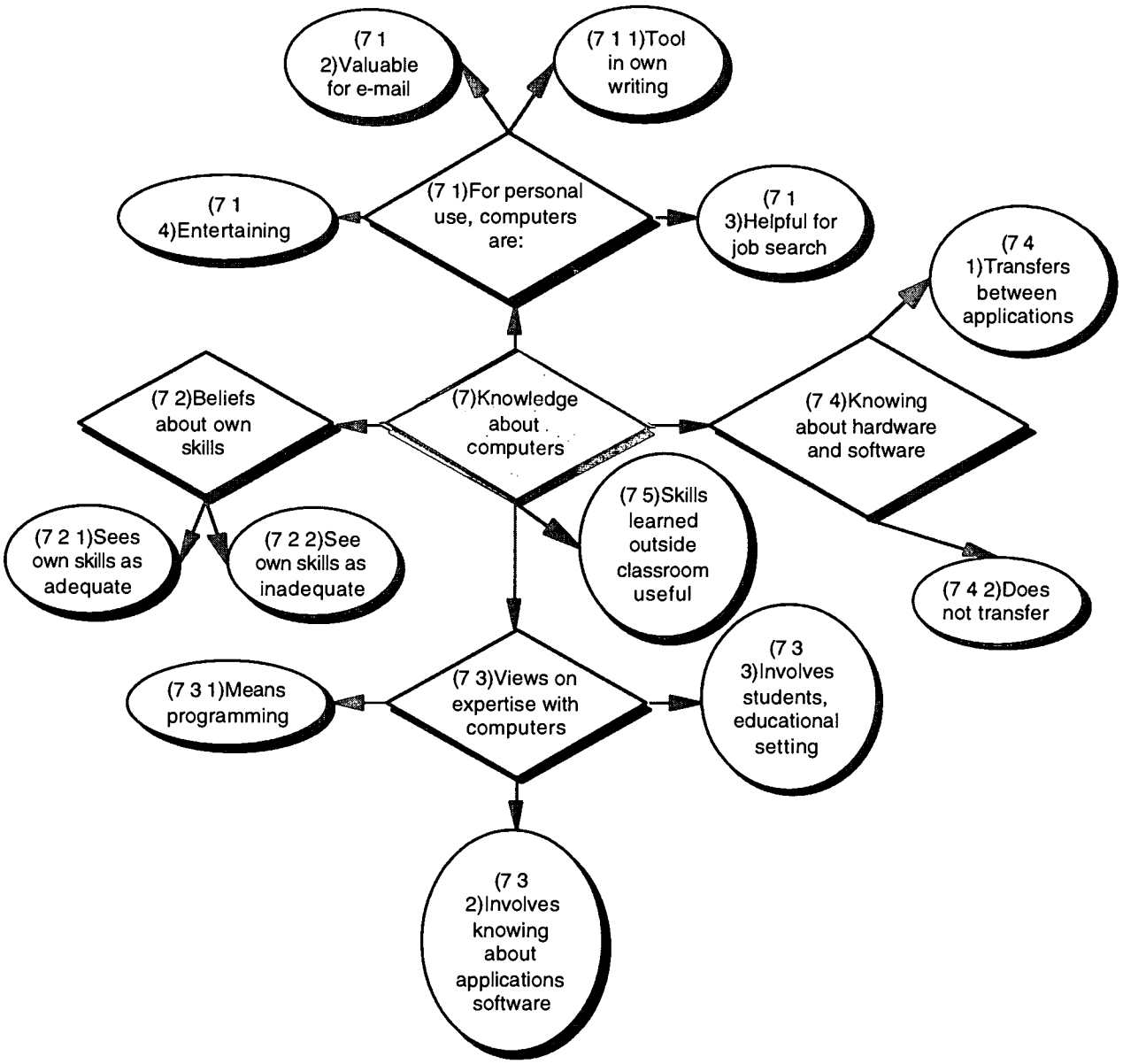
BEST COPY AVAILABLE

Figure 12: Category 6, Learning About Computers (Participants), 10 Codes



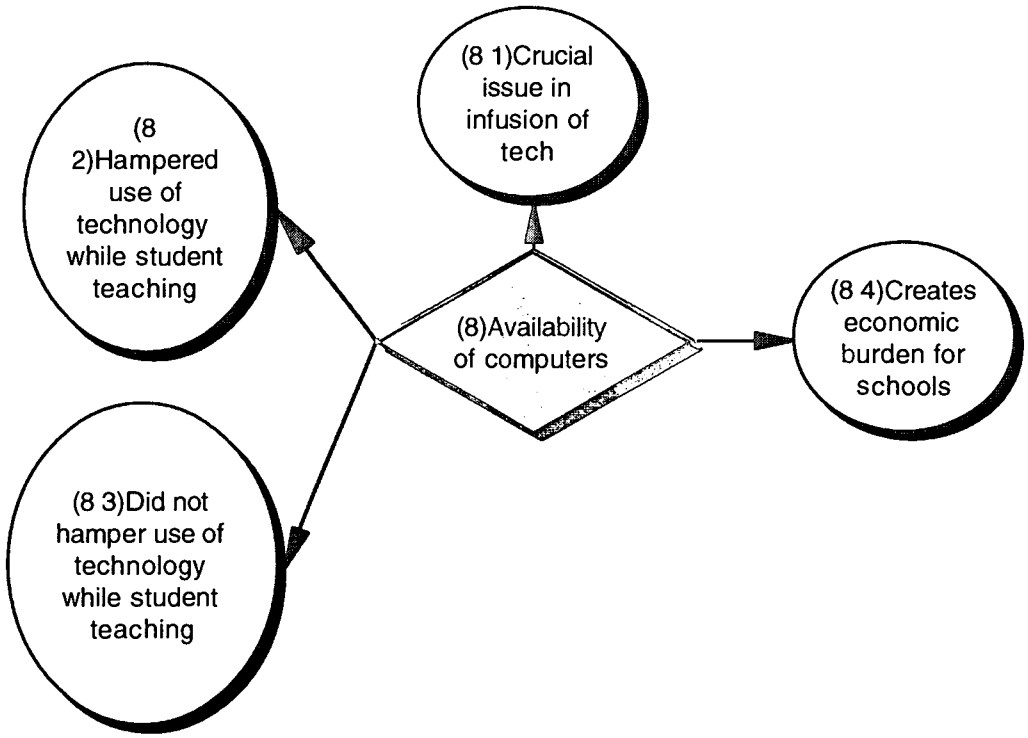
BEST COPY AVAILABLE

Figure 13: Category 7, Knowledge About Computers, 12 Codes



BEST COPY AVAILABLE

Figure 14: Category 8, Availability of Computers, 4 Codes



BEST COPY AVAILABLE

Appendix B: Coding Scheme for Study 2

Coding by student:

- Jorge
- Estella
- Alicia
- Manuel

Coding by theme:

- Kid life
  - ◊ Testing limits
  - ◊ Student-student email
  - ◊ Tech skill as social capital
  - ◊ Peer networking
  - ◊ Inappropriate
  - ◊ Student surfing
- Family life
  - ◊ Anxiety
  - ◊ Family (non) use of tech
  - ◊ Parent email
  - ◊ Parent participation
  - ◊ Parent-child interaction around NetTV
  - ◊ Parent wishes about NetTV
  - ◊ Student leading

- ◊ Push for parent participation

- School life
  - ◊ Web for Homework
  - ◊ Educational Attitudes
  - ◊ T to SS about NetTV
  - ◊ Other computer use
  - ◊ Email for homework
  - ◊ Privileging NetTV use, skill
  - ◊ Academic use of NetTV
  - ◊ Rationale for NetTV

- Value of NetTV to teacher

TM029847

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

AERA 1999

I. DOCUMENT IDENTIFICATION:

Title: Analytical and methodological issues in the use of qualitative data analysis software: A description of three studies

Author(s): Jon Margerum-Leys, Jeff Kupperman, Kristen Boyle-Heimann

Corporate Source: The University of Michigan

Publication Date: April 22, 1999

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]

Printed Name/Position/Title: Jon Margerum-Leys

Organization/Address: 610 E. University, Rm. 1323, Ann Arbor, MI 48105

Telephone: 734-647-2263

FAX:

E-Mail Address: jmargeru@umich.edu

Date: 4/20/99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com