ETHNOGRAPHIC CONTENT ANALYSIS
OF COUPLE AND THERAPIST PERCEPTIONS
IN A REFLECTING TEAM SETTING*

Thomas Edward Smith
Florida State University

Scott P. Sells
University of Nevada-Las Vegas

Theodore Clevenger
Florida State University

An ethnographic content analysis was used to examine couple and therapist perspectives about the use and value of reflecting team practice. Postsession ethnographic interviews from both couples and therapists were examined for the frequency of themes in seven categories that emerged from a previous ethnographic study of reflecting teams (Sells, Smith, Coe, Yoshioka, & Robbins, 1994). The study demonstrated that quantitative numerical data and qualitative narrative data can examine the same phenomenon from multiple perspectives and allow for greater accuracy and stability in study findings. Ethnographic content analysis is briefly contrasted with conventional modes of quantitative content analysis to illustrate its usefulness and rationale for discovering emergent patterns, themes, emphases, and process using both inductive and deductive methods of inquiry.

Since their introduction by Tom Andersen in 1987, reflecting teams or reflecting process have become a popular topic among family therapists in manuscripts, workshops, and practice. However, reflecting team literature has been primarily dominated by theoretical discussions and case studies from the clinician's perspective (Andersen, 1987; Miller & Lax, 1988). Within these studies, many differences exist about the use and process of reflecting team practice. Many clinicians argue that a reflecting team is useful when a family system is "stuck" or needs new information (Andersen, 1987; Kassis & Matthews, 1987). Some clinicians have stated that they prefer the use of the reflecting team to begin therapy, and others have argued for its use in later sessions to open up additional discussion topics (Davidson,
Contraindicated use has been suggested when the problem is of low intensity (Kassis & Matthews, 1987) or when therapy sessions are continuing in a useful manner without the reflecting team present (Smith, Winton, & Yoshioka, 1992; Smith, Yoshioka, & Winton, 1993). The process of change in reflecting team practice has been discussed by Andersen as the family’s ability to hear the same problem in a slightly different fashion. Prest, Darden, and Keller (1990) describe this process as a unique “fly on the wall” phenomenon in which someone experiences him or herself being talked about (while observing the reflecting team) without being part of the discussion (i.e., being at a meta-level to the process).

Although these case studies are helpful to practitioners in forming hypotheses about the success or failure of cases, these studies are not explicitly tied to theory. Important conceptual differences exist about the use and process of reflecting team practice (Jenkins, 1992). This is a significant omission in a treatment model because unclear concepts and assumptions result in families suffering from untestable treatment modalities (Wynne, 1988). Testing treatment models is necessary for a reflexive practice mindset that will prevent dogmatic adherence to faddish practice techniques. Without grounding in data, any theory generated will be speculative, hence ineffective (Strauss, 1987). Research is therefore needed to develop a theory base that would guide the practice and understanding of a reflecting team approach to family therapy.

A first step in this process was ethnographic studies by Smith et al. (1992), Smith et al. (1993), and Sells et al. (1994) in which a rich description of reflecting team practice was created from couple and therapist interviews over a 2-year period. When little is known about a particular phenomenon, a qualitative analysis of data is an optimal beginning to discover theory and generate hypotheses that can be tested empirically (Moon, Dillon, & Sprenkle, 1990; Strauss, 1987). An ethnography is qualitative in character and an appropriate first step for the discovery and development of “grounded theory” that emphasizes the need to convert diverse material (i.e., interviews, field notes, census statistics, questionnaire answers) into theory that characterizes the central phenomenon studied (Glaser & Strauss, 1967; Strauss, 1987). From this inductive analysis, assertions are made and hypotheses generated as a guide to quantitatively focused studies (Moon et al., 1990; Strauss, 1987). As noted by Strauss (1987), both qualitative and quantitative modes of inquiry are absolutely necessary to develop theory. When focused on the same issue, qualitative and quantitative modes of analysis can assess the robustness or stability of research findings (Firestone, 1987; Jick, 1979). In this sense, both methods may be used not only to examine the same phenomenon from multiple perspectives but also to enrich our understanding by allowing for new or deeper dimensions to emerge (Jick, 1979).

The terms *qualitative* and *quantitative*, however, do not adequately describe studies that bridge both methodologies. A better distinction is found in comparing emic and etic perspectives. An emic perspective includes “... accounts, descriptions, and analyses expressed in terms of the conceptual schemes and categories regarded as meaningful and appropriate by the native members of the culture. ...” (Lett, 1990, p. 130). This distinction is important because it suggests that not all qualitative research designs are able to capture the world view of informants. Some qualitative methodologies, content analysis and its cousin conversation analysis, traditionally are etic in their use and are not suitable for understanding informants from their own perspective. Etic perspectives represent the researcher’s beliefs and values (Headland, Pike, & Harris, 1990). For studies to capture the complexity of clinical processes, they must be able to capture both emic and etic perspectives.
Purpose of the Study

In this study, an ethnographic content analysis was used to refine and increase our understanding of theoretical relationships and categories that were developed in the Sells et al. (1994) ethnographic domain analyses, and to refine the description of emergent patterns, emphases, and themes. Using this methodology, categories of description that emerged from the Sells et al. domain analysis were combined with the analytic deductive capabilities of content analysis to investigate the frequency of themes that tap into the underlying (i.e., latent) meanings within transcribed interviews. At the conclusion of the Sells et al. study, two assertions emerged. The first was that content categories revealed that the couples’ perceptions about reflecting teams significantly differed from those of therapists. Specifically, therapists appeared more concerned with differential uses of reflecting team practice and how to increase effectiveness. Couples, however, appeared more concerned with the benefits of the team, gender issues, spatial separateness, and how they could hear their problems differently. The second was that the sequences of communication that allowed a couple to hear problems differently were closely associated with the spatial distance or boundary between themselves and the client-index therapist system. An examination of the domain analysis led to the following research questions:

1. Do therapists’ and couples’ descriptions about the use and process of reflecting team practice differ significantly from each other? If so, which categories appear to be differentially endorsed?
2. Are couples’ self-reports about the process of hearing a problem differently closely associated with their discussions of spatial distance between themselves and the therapists on the reflecting team?

Answers to these questions will lead to a verification of earlier assumptions and/or a discovery and development of new patterns or themes. Combining both qualitative and quantitative procedures on the text provides “a detailed microscopic examination of the data in order to better capture the complexity of the phenomena under study and the grounding of the data into theory” (Strauss, 1987, p. 20).

Qualitative/Quantitative Research and Ethnographic Content Analysis

Within the family therapy field, there has increasingly been a call for incorporating more qualitative research methods (Gale & Newfield, 1992; Moon et al., 1990; Newfield, Kuehl, Joanning, & Quinn, 1991) and multimethod family therapy research models that apply both qualitative and quantitative approaches (Atkinson, 1992; Boss, 1992; Moon & Sprenkle, 1992). As noted by Atkinson (1992):

Qualitative methods explore and quantitative methods legitimize. The finding that qualitative methods generate must be empirically verified through rigorous methods of conventional (quantitative) research. Qualitative research is, however, a vital part of this process. Without it, researchers would remain too far removed from the phenomenon of interest to be able to generate creative hypotheses for testing. (p. 15)

Ethnographic content analysis (ECA) was first proposed by Altheide (1985, 1987) to incorporate qualitative methods of ethnography with quantitative methods of content analysis within a single study. The researcher begins with a discovery-oriented ethnographic design to generate descriptive categories and theoretical concepts directly from detailed descriptions from a group of people (i.e., informants) from within a particular setting of interest (Atkinson & Hammersley, 1989). These detailed descriptions are obtained through open-ended, exploratory interviews and observations that generate core categories (i.e., emergent
themes) across all the interviews and observations collected in the study. Within ECA, concept development, sampling, data collection, data coding, data analysis, and interpretation are reflexive (Altheide, 1987; Plummer, 1983).

Once the researcher has derived categories and a coding system from the ethnographic study data, a quantitative content analysis design provides a set of procedures to code the categories systematically with reliability checks to analyze, validate, and report the results (Altheide, 1987). The aim is to be systematic and analytic but to avoid dogmatic procedures. Although categories that emerged previously from ethnographic analysis can guide the study, other categories are allowed and expected to emerge through a content analysis of narrative text. According to Altheide (1987), ECA is an informal version of the constant comparison method in which newly collected information is compared to already existing information (cf. Gale & Newfield, 1992; Strauss, 1987). Newly collected information that is discrepant from extant information necessitates revision of existing categories. The traditional content analytic approach differs from ethnographic content analysis in that traditionally, categories are defined a priori and the numeric frequency of categories is analyzed without regard to the underlying meaning of the speaker (cf. Berelson, 1966; McCormack, 1982). ECA, however, draws on and collects both numerical and narrative data, rather than forcing the latter into arbitrary, predefined categories of traditional content analysis. In short, while categories can still be connected and put into categories, ECA also allows for a narrative description to interpret results better. In this way, theoretical relationships are verified and new concepts that emerge during the research process are discovered. Hence, data analysis is both statistical and textual (Altheide, 1987).

Reliability

Reliability and validity of claims arising from ethnographic content analysis are supported through intercoder reliability and informant verification. The former is a quantitative content analytic procedure whereby categories are tested for reproducibility when coders, unaware of the study purpose, analyze the same text and produce comparable results (Weber, 1990). The latter refers to the extent that agreement is achieved among and confirmed by the people who produced the written documents (LeCompte & Goetz, 1982). ECA incorporates both intercoder reliability and informant verification in order to check, supplement, and confirm prior theoretical claims. Concept categories generated must be confirmed by the individuals who produced them and defined precisely enough to ensure reproducibility and consistency of shared understanding and meaning by multiple observers (LeCompte & Goetz, 1982; Weber, 1990). High reliability is critical to meaningful ECA because it (reliability) measures the consistency of different researchers’ private understanding of the phenomena studied (Krippendorff, 1980; Weber, 1990).

Validity

Validity of these claims is supported by comparing numerical data from content analysis with ethnographic data of the same construct. Hence, validity entails generalizing construct categories across measures or methods (Weber, 1990). Validity is supported by the use of examples from the text itself to demonstrate claims, agreement between researchers and informants (i.e., people who produced the text), and high reliability of multiple coders.

In sum, ethnographic content analysis lends itself to both qualitative and quantitative research goals and combines what are usually considered to be antithetical modes of analysis. With ECA, couples and therapists teach researchers about indigenous meaning worlds.
Categories and hypotheses are generated from theory-rich descriptions. Content analytic procedures are employed in a reflexive and highly circular process among concept development, data coding, analysis, and interpretation. Because ethnography and content analytic procedures interface and complement one another, similarities and differences among text are revealed that would be difficult, if not impossible, to detect otherwise.

METHODS AND PROCEDURES

Sample
Eleven couples and five therapists were interviewed at least twice over a 4-month period concerning their reactions to and perceptions of reflecting team practice. Participants were selected using an opportunistic sampling strategy. Opportunistic sampling is similar to convenience sampling and requires researchers to select available participants who have information relevant to the topic of inquiry (Honigman, 1970). Opportunistic sampling was well-suited for this study because generalization to the population was not the research goal. Instead, the goal was to generate vivid descriptions of reflecting team practice and generate assertions that lead to theory (Babbie, 1973; Yin, 1989).

Couples had been married at least 1 year. Wives ranged in age from 21 to 49 years (mean = 35); husbands ranged in age from 24 to 51 years (mean = 37). Couples who requested marital counseling services at the university Marriage and Family Clinic were informed about the research project and were given a choice to participate in the study and in reflecting team sessions. Therapists included four doctoral students—two males and two females—and a faculty supervisor from a marriage and family therapy program. Doctoral students’ experiences ranged from 1 to 6 years. The male faculty supervisor had 12 years experience, including 7 years of experience in using one-way mirrors in systemic practice.

Interview Procedures
Each spouse was interviewed immediately following the first reflecting team session and once again several weeks later to verify and expand on statements from the first interview. Initial interviews averaged 30 minutes to 2 hours in length. Subsequent interviews ranged from 1 to 2 hours in length or until the informants reported that they had no more to say in the interview. Interviews with spouses were dialogical in nature; although spouses were aware that their conversations were being recorded, the interviews had the flavor of a debriefing in which multiple topics were addressed in a fluid conversation. Because of confidentiality concerns, the spouses were not interviewed in a group format.

Each therapist was interviewed three times in a group discussion format. This allowed for the creation of theoretical or analytical memos (Bogdan & Biklen, 1982; Strauss, 1987). Therapists’ memos were reflections that focused on emerging themes that might be present. Memos in this study were generated after every two reflecting team sessions through collaborative brainstorming that used an open discussion group format. Therapists also completed field notes immediately following each session. Therapists did not speak to each other about sessions until after they had dictated field notes into an audiotape recorder (for a discussion of this process, see Sells et al. [1994]). All interviews and field notes were recorded using an audiotape recorder and transcribed by research assistants. Content analytic procedures are performed directly on participants’ transcripts. This yields nonreactive measurements because coding transcripts is unobtrusive; the message senders (i.e., the couples and therapists) are not being analyzed directly but indirectly through their transcripts. One important goal
in this study was to maximize the number of replicable research procedures. Thus, one important advantage of our content analytic procedures was to increase the likelihood of replicability. However, the overall study was a balance between emic and etic perspectives; although our coding and content analytic procedures are replicable, the initial category development was unique to the clients and therapists of our setting. Hence, it is likely that another setting with other clients may well produce different results.

Developing Coding Categories in Ethnographic Content Analysis

Content analysis is essentially a coding operation (Rubin & Babbie, 1989). The many words, sentences, or paragraphs of the text are classified into many fewer content categories (Weber, 1990). Categories provide the structure for grouping the units of analysis into the same conceptual units that have similar meaning (USGAO, 1989). Their importance is emphasized by Berelson (1952) when he cautions that: “Content analysis stands or falls by its categories. Particular studies have been productive to the extent that the categories were clearly formulated and well adapted to the problem” (p. 92).

A dilemma faced in the field of conventional content analysis (Rubin & Babbie, 1989; Stemple & Westley, 1990; Weber, 1990) has been the complexity and difficulty of category construction. These problems stem from the ambiguity of category definitions and words or sentences that are to be assigned to categories (Weber, 1990). Historically, content analysis researchers were advised to use a category system that had already been in use, or code manifest surface content (i.e., frequency of words) as it appears, rather than as researchers believe it was intended (Berelson, 1952; Stemple & Westley, 1990). However, accepting content purely at its face value or manifest level may lead to distortions. Simply to count the number of times a word is used may fail to project the underlying or latent content of the communication (Krippendorff, 1980; Weber, 1990).

Herein lie both the strengths and weaknesses of content analysis methodology. A key strength is that, once the researcher has categories and a coding system, content analysis provides a set of procedures to code categories systematically with reliability checks to analyze and interpret the information and to validate and report the results. As Weber (1990) notes, quantifying latent characteristics of text often results in more precise comparisons of textual items and allows detection of similarities and differences that would be difficult without such quantification.

Its weakness is the lack of agreement and even absence of systemic and detailed information on how to construct categories. Suggestions are made that the categories are to be mutually exclusive, exhaustive, and independent (USGAO, 1989), but the construction itself is left to the subjective interpretation of the researcher. As Weber (1990) states:

The coding and quantitative technique in content analysis of latent characteristics of text are criticized because they do not make much use of the qualitative syntactic and semantic information in each sentence. Instead, categories concerning latent characteristics of text are developed in the absence of accompanying detailed semantic and syntactic information. (pp. 74-75)

Category construction. A solution to this dilemma can be found within the framework of ethnographic content analysis. Unlike traditional content analysis in which categories are predefined, ethnographic content analysis is oriented to construct categories qualitatively in order to develop analytical constructs appropriate to this study (Altheide, 1987). In the Sells et al. (1994) study, semantic and syntactical information within each sentence was examined through an ethnographic domain analysis based on Spradley’s (1979) Developmental
Research Sequence Model (DRS). In a domain analysis, narrative is coded into phrases or words that have a semantic relationship with their cover terms. For example, one couple stated, “I listen to them [the reflecting team] because I can’t object, I’m forced to listen, and I have to digest it.” Using a domain analysis, “forced to listen,” “can’t object,” and “have to digest it” are all included terms that cause (semantic relationship of cause-effect) this couple to “listen to the reflecting team” (cover term). Therefore, the emerging domain for this group of sentences would be “causes of having the reflecting team.” Constructing a list of domains like this one within each interview and set of field notes is a first step in capturing the underlying or latent meaning that therapists and couples use to interpret their experiences within a reflecting team context.

In domain analysis, domains and coding categories emerge concurrently as the same domains are repeatedly presented by couples and therapists. By definition, domains have one or more features in common and cannot overlap. They can however be subsumed into larger, more inclusive domains. Included terms within domains provide connotative meaning; included terms provide examples of a domain and allow readers to develop an appreciation of the meaning of a domain. By making comparisons among these domains, a qualitative analyst is forced into confronting similarities, differences, and degrees of consistency of meaning among domains. The process of subsuming domains into larger domains and sharply differentiating different domains results in a series of mutually exclusive categories. Such categories form a grounded, empirically derived coding schema that is ideally suited for content analysis. In many ways, this process is similar to the open-coding procedures described by Glaser and Strauss (1967) and Strauss (1987). Both examine the linguistic features of communication by simultaneously coding and analyzing the data in order to develop concepts that result in mutually exclusive coding categories.

Rigorous domain analyses systematize the qualitative process of coding large blocks of text which leads to categories that are amenable to the statistical calculations within content analysis. This analysis is a necessary process when symbolic communicators in the form of words carry information about the phenomena outside their physical manifestation (Krippendorff, 1980). It is here that content analysis methods combine what are usually thought to be antithetical modes of analysis. Categories are generated through a qualitative methodology of domain analysis and subsequently analyzed with quantification methods of content analysis. In the following statements Weber (1990) notes how this is useful to the researcher:

If one has observables such as texts[,] why bother with quantification and latent-indicator models at all? There are several reasons. First, counting generates results that allow for more precise comparisons among texts. Second, we want to know how much more (or less) attention is devoted to some texts than to others. Third, quantitative analytical procedures often reveal similarities and differences among text that would be difficult, if not impossible, to detect otherwise. (p. 74)

The above discussion, however, does not do justice to dynamic features of data collection and reduction present in ethnographic content analysis. As Altheide (1987) wrote:

Ethnographic content analysis consists of a reflexive movement between concept development, sampling, data collection, data coding, data analysis, and interpretation. Although categories initially guide the study, others are allowed and expected to emerge throughout the study. (p. 68)

In this ECA study, initial categories from the Sells et al. (1994) study were used to guide the study, but other categories emerged as more data were collected and analyzed in interviews with both therapists and couples. One category, “Index Therapists’ Use of the Reflecting
"Team" emerged as a result of content analysis; this category was added when coders examined and reexamined the text. In addition, one category from the Sells et al. (1994) study, "Spatial/Process" was divided into two distinct categories, "Spatial Separateness" and "Process of Hearing," as a result of the additional coding of the text. Thus, the ethnographic content analysis was effective in refining and discovering new categories of meaning. Following are the final seven code categories and definitions that began with the domain analysis in the Sells et al. (1994) study and were refined in this study:

1. **Benefits or Reasons to Have a Reflecting Team.** Statements made by couples or therapists to describe things about the reflecting team that they liked or instances when the team was helpful in some way.

2. **The Impact of Therapists’ Gender.** Statements made by either couples or therapists that describe why it was beneficial to have both male and female therapists on a reflecting team.

3. **Index Therapists’ Use of the Reflecting Team.** Statements made by therapists and couples that describe how the index therapist used the team or received its assistance in his or her counseling with the couple.

4. **Recommended Use of the Reflecting Team.** Statements made by therapists (i.e., team members) and couples that describe how the team should be used to do the most good or to be optimally effective.

5. **Specific Times When a Reflecting Team Was Not Useful.** Statements made by couples and therapists that describe negative aspects of the team or specific instances when a reflecting team was not helpful or should not have been used.

6. **Times When Couples Were Able to Hear What the Reflecting Team Therapists Said.** Statements made by couples or therapists that indicated the process of hearing by which couples were able to sit back and hear what the team was saying to them.

7. **Examples and Effects of Spatial Separateness Between the Therapists on the Reflecting Team and the Couple.** Statements made by the couple or therapist that indicated the effects of having a distance or separateness within the therapy room between the couple and the reflecting team.

**Refining and Implementing the Coding Category System**

In content analysis, applying a coding scheme consists of the following steps: (a) encoding text, (b) test coding a sample of text, (c) assessing reliability on the sample text, (d) revising coding rules, and (e) assessing final reliability. Because a central problem in content analysis originates in the consistency or reliability of the coding scheme employed, it is crucial that it be carefully chosen or developed (Weber, 1990). Categories must be reliable and reproducible in the sense that multiple coders analyzing the same text produce results comparable to those of the researcher. Disparate coding may result from cognitive differences among coders, ambiguous coding instructions, or ambiguous word meanings. High reliability is a minimum standard for content analysis (Krippendorff, 1980; Rubin & Babbie, 1989; Weber, 1990). High reliability of codes helps establish whether data can provide a trustworthy basis for drawing inferences, making recommendations, or supporting decisions (Krippendorff, 1980). The iterative process of subjecting a sample of text to reliability tests and revising the coding schema before applying it to the entire transcript ensures that time and effort is not wasted on nonreproducible findings.

**Text encoding.** In this study, coders were trained in a coding schema (i.e., seven categories) listed above. A stratified sample based on the date of interviews and field notes was
Jeff: I like input, especially in this situation as much as possible. And if you came to me and said hey man I want to learn how to be a drummer, I would take you to a point and then I would introduce you to some other drummers. That other input is critical. I mean I may cover areas that are just real cool, but I may not have thought of this. Not to say that I’m a lesser drummer.

Jack: I think that’s a pretty fair assessment. I think that any interference, because when they’re back there, you’re wondering what they’re thinking, what they’re doing. But with them sitting back there it brings it more to light so you can kind of sit back and hear what they’re thinking or saying. I don’t know, it just made it seem more open, I don’t like the mystery behind the mirror.

Amy: You can’t project to the other person. This is not so personal, when you’re sitting back listening to other people talk about you it’s like you’re outside the problem and you can see it differently. It made me more objective. The longer I had to sit there I understood more. When we first walked into one. And its sort of like being a bug on a wall. So far I haven’t wanted to say anything while they’re talking anyway. But if I could, I would wait until she put the mirror back down and said, well here’s what they had to say, what do you think?

Billy: Because I can hear all opinions instead of just the one. And its sort of like being a bug on a wall. So far I haven’t wanted to say anything while they’re talking anyway. But if I could, I would wait until she put the mirror back down and said, well here’s what they had to say, what do you think?

Code Translations

1 = Benefits of RT
2 = Male/Female Therapists on RT
3 = Couples’ Process of Hearing RT
4 = Ways Index Therapist Uses RT
5 = Specific Times RT Useful
6 = Specific Times RT Not Useful
7 = Spatial Separateness.

encoded by multiple coders and tested for reliability before the entire transcript was coded.
of text can be located, copied, labeled, and directed to a screen, printer, or text file. Codes are then electronically attached to the text without altering the document and codes and code frequencies compiled from a single sample or entire database.

Other programs (The Ethnograph, Tap, Textbase, Alpha) contain similar features, but Qualpro computes and reports intercoder reliability after adjusting for chance results. Other content-analytic software programs omit the ability to compute intercoder reliability and require manual computation of reliability estimates between coders; in many instances researchers fail to adjust statistically for chance findings (Stemple & Westley, 1990). In addition to reporting reliabilities between coders, Qualpro generates a Table of Agreements and a Confusion Matrix that assist the researcher in identifying exactly where sources of unreliable judgment between coders exist. The Table of Agreements is a complete list of all the coding units in a content sample showing the content category to which each was assigned by each coder and indicating for each unit whether the coders agreed or disagreed on that unit. It is used for calculating percentage of agreement and identifying agreement-disagreement on individual content units. The Confusion Matrix is a two-dimensional contingency table, similar to Krippendorff's (1980) Coincidence Matrix, showing for two coders the frequency with which each coding category co-occurs (is confused with) each other category in a content sample. The set of content categories is arrayed vertically for one coder and horizontally for the other. The diagonal frequencies represent agreements, and all off-diagonal frequencies are disagreements. The Confusion Matrix is used to identify recurrent patterns of disagreement between the two coders.

In general, text database managers help avoid the tedious and time-consuming process of the “cut up and sort” or file card systems. In early content analysis, researchers took pairs of scissors, sat in the middle of a floor, cut up the hundreds of pages of transcripts, and placed them in manila folders, each labeled with one code (Bogdan & Biklen, 1982). The sample of text in Figure 1 presents how the transcripts were coded. Each code was then entered into Qualpro for subsequent analysis.

Test coding on stratified sample of text. Typically, testing on a sample of text reveals ambiguities in coding rules the clarification of which leads to significant revisions of the coding scheme (Krippendorff, 1980; Weber, 1990). Typically, two or three reliability trials on a sample of text are conducted before researchers feel confident to proceed with establishing a benchmark reliability estimate and generating the remainder of the text for analysis (Krippendorff, 1980).

In selecting a test sample for this study, stratified random sampling was used to ensure representativeness and decrease probability of sampling error (Babbie, 1973; Rubin & Babbie, 1989). Distinctions between dates were known to exist based on an earlier study (Sells et al., 1994). Because the earlier study had used Spradley’s Developmental Research Sequence Model (Spradley, 1979), questions over time had become less general and descriptive and more specific and structural. Thus, the stratified random sampling was necessary to ensure that coders would be able reliably to code the statements that resulted from questions that had changed over time. Interviews and field notes were stratified by the date of the interview and then paragraphs of text within each interview were randomly selected. Table 1 in Sells et al. (1994) shows how questions changed depending on their iteration.

Reliability of coding on stratified sample. Interrater reliability was assessed using a Scott's phi (Scott, 1955) that varies from 0 to 1 and corrects for chance agreement. Two graduate students in social work were trained as coders and then independently coded text that the authors had previously coded.Coder 1’s percentage agreement with the researchers was .71.
Coder 2’s percentage of agreement was .69. With no previous studies of a similar category system, acceptable reliabilities were based on a criterion developed by Krippendorff (1980). He recommended that a reliability above .80 provides a trustworthy base for drawing inferences and supporting category definitions.

An examination of Qualpro’s Confusion Matrix and Table of Agreements revealed that the definitions of process of hearing and spatial separateness accounted for the largest percentage of disagreement. Confusion of this kind usually results from ambiguous definitions or coders’ inability to understand the meaning of that category (Krippendorff, 1980). When these categories are removed, checked, revised, or recoded, data reliability improves (Weber, 1990).

Revising coding rules. When serious errors are discovered, coding rules are typically revised (Rubin & Babbie, 1989; Stemple & Westley, 1990). One category, “Index Therapists’ Use of the Reflecting Team,” emerged as a result of content analysis; this category was added when coders examined and reexamined the text. Index therapists often discussed how reflecting teams assisted them in resolving impasses. Thus, this category was added to the original set of categories from Sells et al. (1994). Problems also resulted from ambiguities in the definitions of the categories, “Process of Hearing” and “Spatial Separateness.” Collaborative brainstorming with coders revealed that couples occasionally spoke about process of hearing and spatial separateness in the same coding unit (i.e., sentence or paragraph). Such coterminous events resulted in discrepant codes since the original codes were designed to be mutually exclusive. To delve further into this interchanging of codes, we spoke briefly with several couples by phone to gain further understanding of the differences. Based on couples’ and coders’ input, revisions of these two categories were then implemented to allow for more consistent and accurate coding. In summary, one new category (“Index Therapists’ Use of the Reflecting Teams”) was added and one category from the earlier study was divided into two categories (“Process of Hearing” and “Spatial Separateness”).

Assessing final reliability. Another stratified sample was taken and interrater reliability equaled .81 for Coder 1 and .84 for Coder 2. When high reliability is achieved, the coding rules typically can be applied to the entire text (Krippendorff, 1980; Weber, 1990). In this study, achievement of high reliability estimates gave us confidence in our category definitions, allowing coders to proceed with the subsequent coding of the entire text.

RESULTS

Data generated by the coding system were analyzed using quantitative frequency counts of the seven categories and qualitative narrative data from the text. Statistical findings regarding the latent meaning of the interviews were validated by references to the text (Altheide, 1987; Weber, 1990). In conventional content analysis this does not occur; however, in an ethnographic analysis, narrative and numeric data are presented concurrently (cf. Altheide, 1985, 1987). Frequency counts of categories were analyzed using a chi-square statistic. This statistical procedure was used to test the comparability of therapist and couple perceptions across all seven categories. Differences between couples and therapists within each category were calculated proportionally and compared using a chi-square to determine their significance.

A phi coefficient contingency analysis was used to measure the correlation of the categories “Spatial Separateness” and “Process of Hearing” co-occurring together. To test whether the observed value of the phi coefficient indicated a significant association between
spatial separateness and process of hearing, a chi-square test was computed. Text that contained the codes “Spatial Separateness” and “Process of Hearing” were examined for the highest frequency of words. Ordered word-frequency lists were augmented with the text in which each word appeared. This data base was then used to study detailed differences and similarities in semantic word usage within the “Spatial Separateness” and “Process of Hearing” categories.

Through the use of Qualpro (Blackman, 1992), all coded sections of the text were located and directed to the printer and the text file. This allowed us quickly to find narrative examples of each code category to examine the study’s statistical conclusions. In this way, a qualitative ethnographic analysis supported a quantitative content analysis of the data.

Research Questions and Subsequent Results

1. Do therapist and couple perceptions about the use and process of reflecting teams practice differ significantly from each other? A two-independent sample chi-square test was performed on the observed and expected frequencies between the two groups ($\chi^2 = 122.52$, $p < .001$, $df = 6$, $N = 465$); formulae for calculating the chi-square statistic and expected frequencies are available from Siegel and Castellan (1988). Table 1 shows which categories couples and therapists considered to be least and most important.

Table 1

<table>
<thead>
<tr>
<th>Categories</th>
<th>Group</th>
<th>Benefits</th>
<th>Gender</th>
<th>Process</th>
<th>Ways</th>
<th>Rec Use</th>
<th>RT Not Use</th>
<th>Spatial</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>Couples</td>
<td>69</td>
<td>26</td>
<td>48</td>
<td>22</td>
<td>39</td>
<td>21</td>
<td>49</td>
<td>274</td>
</tr>
<tr>
<td>Expected</td>
<td>47.1</td>
<td>25.3</td>
<td>31.8</td>
<td>43</td>
<td>54.8</td>
<td>39.4</td>
<td>32.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentages</td>
<td>25%</td>
<td>9%</td>
<td>17.9%</td>
<td>8%</td>
<td>14.2%</td>
<td>7.7%</td>
<td>17.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>Therapists</td>
<td>11</td>
<td>17</td>
<td>6</td>
<td>51</td>
<td>54</td>
<td>46</td>
<td>6</td>
<td>191</td>
</tr>
<tr>
<td>Expected</td>
<td>32.8</td>
<td>17.6</td>
<td>22.2</td>
<td>29.9</td>
<td>38.2</td>
<td>27.5</td>
<td>22.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentages</td>
<td>6%</td>
<td>9%</td>
<td>3%</td>
<td>26.7%</td>
<td>28.3%</td>
<td>24%</td>
<td>3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>43</td>
<td>54</td>
<td>73</td>
<td>93</td>
<td>67</td>
<td>55</td>
<td></td>
<td>465</td>
</tr>
</tbody>
</table>

Note. Summary of a two sample chi-square performed on couple and therapist category frequencies ($\chi^2 = 122.52$, $p < .001$, $df = 6$, $N = 465$). Percentages represent the proportion of observed frequencies for couples and therapists respectively.

Benefits = Benefits of RT
Gender = Male/Female Therapists on RT
Process = Couples' Process of Hearing RT
Ways = Ways That Index Therapist Uses RT
Rec Use = Specific Times RT Useful
RT Not Use = Specific Times RT Not Useful
Spatial = Spatial Separateness
Not surprisingly, couples’ and clinicians’ interviews revealed significant differences between the coded categories, with a notable exception within the category of gender. A contingency table analysis yielded a nonsignificant chi-square statistic ($\chi^2 = 1.03, p > .05, df = 1, N = 495$). Unlike the other six categories, therapist and couple perceptions with regards to the category of gender were comparable in their emphasis. Therapists devoted about 9% of their comments to gender; couples also devoted 9%.

With regard to the category of gender, the text supported the statistical conclusions. Overall, therapists and couples talked about gender with the same frequency. Also, a close examination of the text revealed several conceptual similarities. Couples frequently commented that there existed inherent differences between men and women, or that men and women have different points of view. They believed that it was beneficial to have both men and women on the reflecting team in order to feel more comfortable and obtain different perspectives. Both couples and therapists reported a sense of equality or balance by having both male and female team members present.

**Couple #11:** Sometimes a man and a woman will have a different perspective on life, and it helps to get both ideas by having a man and a woman in the room.

**Couple #7:** I think having a male and female made it fairer. If you had two women on there it couldn’t have been objective because the guy would have felt like he was going to get ganged up on.

Therapists also revealed this sense of equality or balance by having both a male and female perspective.

**Field note #6:** Reflecting team’s ability to have gender splits was noticeable during this session. Because the index therapist was taking a flagrant position in favor of the woman, I felt drawn to support the man. I think J came out pretty heavily for the man and that helped balance out the therapist’s comments.

**Ther. Int. #1:** I think it’s also important to ensure different voices present as in male and female voices. No matter how good the therapist is he or she is not going to be able to get both genders in.

Regarding the other categories, couple and therapist emphases were quite different. Therapists appeared more concerned with and placed much more emphasis on how the reflecting team should or should not be used and how the index therapist can utilize the reflecting team. In this respect, practice within a reflecting team was regarded more as a method to intervene with clients than as a process or series of developmental sequences that were beneficial or elicited change.

**Ther. Int. #2:** I feel that the reflecting team can say what I’m thinking, to raise the intensity or say things that are more controversial or if I want things in the session to be said that are more controversial then I’ll go to the team.

**Ther. Int. #3:** When there’s no problem, the team should not be used. If the client has in a large part resolved what he or she came in for, the therapist may not feel a need to go to the team and the team will be stuck there doing nothing.

**Ther. Int. #1:** I use the team when I’m intensely involved with the couple and I’m not sure what direction to take. Then I will take a break and listen to the team’s thoughts to get direction and piece things together so I have some kind of map.

Couples, however, were much more concerned with all the benefits a reflecting team format offered. As discussed previously, a primary benefit was the process in which couples were able to sit back and hear or see a particular problem from a different perspective. Couples also appreciated that the team offered three or four other perspectives and that they...
came out from behind the one-way mirror. This allowed them to see team members’ faces while listening to their opinions.

Couple #1: I have knowledge of the fact that these are the same people that have been behind the glass. I find it far less intimidating. I can see that they are not judging me. They’re not interested in what I look like; they’re interested in what I am saying. This gave me confidence.

Couple #3: It was good for me. It helped to give me some insight into different ideas and things I may not have thought of. To actually see what’s happening... that was so important to me.

2. Are couples’ self-reports about the process of hearing the problem differently closely associated with their discussions concerning spatial separateness? A phi coefficient 2 X 2 contingency table analysis (Siegel & Castellan, 1988) yielded a correlation of .78. Using a chi-square test, the correlation of .78 is significant ($\chi^2 = 35.1, p < .001, df = 1, n = 274$). These co-occurrences of spatial separateness and process of hearing were supported by many examples throughout the text. Most couples noted that they were aware of this imaginary boundary between themselves and the team. However, what was interesting about this occurrence was that the concept of spatial separateness was significantly correlated with the concept of process of hearing. The distance created by the imaginary mirror or the therapists themselves was the first or concurrent step that the couples perceived in the process of going outside the problem to see or hear it differently.

Couple #4: It’s like there’s the issue but I can’t jump in, so as they are discussing the problem I can sit back, take a break and listen to them dissect it. I get a better idea.

Couple #7: When you’re sitting back listening to other people talk about you, it’s like you’re outside the problem and you can see it differently. It made me more objective. The longer I had to sit there the more I understood.

Concepts or ideas that are closely associated conceptually are frequently related statistically (Krippendorff, 1980). This was shown to be the case and supports the earlier work by Prest et al. (1990), who described this process as a “fly on the wall” phenomenon in which an imaginary boundary allows someone to go to a meta-level to his or her own process of hearing. One couple even used the fly on the wall metaphor in their description of the process of hearing.

Couple #10: I hear all opinions instead of just the one. And it's sort of like being a fly on a wall. It's like the mirror goes up then—I'm hearing it but I can't object. Many times you want to jump in, but the rest of the thoughts you hear kind of clear up whatever you were going to object about. So it makes you listen better.

The concepts of spatial separateness and process of hearing were different conceptually but related in how couples used them to describe the process of hearing reflecting team members. The former indicated statements pertaining to a boundary created by the team and the setting. The latter pertained to a sequence of stages or instances whereby someone heard the problem differently. However, the two concepts overlapped considerably because the use of imaginary boundaries and enforced silence often preceded couples’ ability to hear the problem in a different manner.

Interestingly, therapists only mentioned the occurrence of spatial separateness and process of hearing a total of 12 times (six spatial separateness and six process of hearing) in all the field notes and interviews combined. In many of these instances, therapists did not recognize the concept spatial separateness as contributing to couples hearing the problem differently. Instead, therapists believed that it was a strategy or technique they used as a team member.
to get the couple to hear them constructively.

**Ther. Int. #3:** The most important part of the team process was right when the therapist first talks, because that's when they hear it the most. I think the key to it is using their language. The more you can use their language and affect, the better chance they have of hearing you.

**Field Note #11:** When a reflecting team actually gets into a good, solid specific conversation, it allowed this couple to pick up a frame. By being specific, I think they can understand quicker and can use what is said to create their own new frame of reference to the problem.

**Key word in context.** To understand why a particular code category or categories like “Spatial Separateness” and “Process of Hearing” might show an intriguing relationship, researchers want to know which words appear frequently and how they are actually used (Webber, 1990). The importance of word context is two-fold. First, words that appear frequently within these categories represent what couples considered most important in “Hearing the Reflecting Team Members.” Second, identifying the context in which each of these words appears helps determine and draw attention to the variation or uniformity in word meaning and usage. Detailed semantic information in each sentence gives insight into how couples understand and give meaning to a reflecting team process. Within the spatial/process code file in Qualpro, we examined words with the highest frequency. The highest frequency of words were the following (with the number of occurrences in parentheses): (a) hear, hearing, heard (47); (b) different (28); (c) understand, understood (22); (d) see (36); (e) listen (39); and (f) mirror (22).

A key word in context (KWIC) was then used to show the context in which each word appeared. Each of these words was an important symbol in understanding couples’ description of the concepts spatial separateness and process of hearing. Some examples of the most frequently used words:

1. The verb to hear
   The verb to hear was used consistently in each of its 47 occurrences. The verb was used by couples to describe their experience of hearing the problem differently from their initial expectations. Examples of their reactions included:
   * You can sit back and hear what they’re thinking
   * They’re adding something, you change where you hear it
   * You hear then and act like well you can’t object

2. The word see
   Another finding was the word see. Couples reported two distinct important usages of the word. First, couples could see the problem or situation in a different light:
   * You see things differently from the outside without interacting with them
   * It gave us a chance to see the problem differently
   Second, couples reported that when they could see other people within the room rather than behind the mirror, and see people analyze the situation, it helped them in their ability to analyze the problem differently:
   * It was interesting to me to see what other people think
   * In a way you can see people interpret my problem
   * Seeing whether I’m being understood was so important to me

3. The verb to listen
   The word listen was also significant in its usage. For couples, their ability to listen was attributed to the team’s physical presence in the room and their actions. Couples stated that
this helped them to pay attention and focus more closely on what the team said.

*You have to sit back and just listen
*The team makes you listen more thoroughly
*You’re forced to listen to a complete thought because the team will not let you talk

DISCUSSION

This study has generated a wealth of information regarding both couples’ and therapists’ descriptions of a reflecting team approach to marital therapy. Tentative beliefs that emerged from the Sells et al. (1994) study were examined using an ethnographic content analysis model to assess their stability and discover new and deeper dimensions of reflecting team practice. Use of an ECA research model combined what are usually thought to be antithetical modes of inquiry; the analysis was both statistical and textual. Category construction was not arbitrarily predefined but developed qualitatively in conjunction with the syntactic and semantic information within the text. Quantitative content analysis of a qualitative domain analysis of the couples’ and therapists’ language yielded an ethnographic content analysis.

Using this methodology, similarities and differences among the text were discovered that were unclear in the purely ethnographic study of the therapist and couple interviews (Sells et al., 1994). One of the most intriguing aspects of the study was how couple and therapist perspectives differ. In the previous study, differences were suspected, but their extent could not be judged or quantified. However, through the constant comparison of both a statistical analysis and continued narrative description, emergent themes and patterns became apparent.

Results revealed that for couples the categories “Benefits,” “Process of Hearing,” and “Spatial Separateness” accounted for approximately 61% of all the categories; therapist categories—“Recommended Use,” “Contraindicated Use,” and “Index Therapist”—accounted for approximately 78% of total categories. Couples repeatedly stressed things about the reflecting team that they liked and the process whereby they could hear their problems differently. Therapists apparently were more concerned with the strategies of therapy outcome than with therapy process. Couples’ attention to process was most clearly revealed directly following a reflecting team interview in which the couple reported feeling “support” and “more input” as a benefit, and therapists independently reported support as “not a good use of the team.” The different perceptions suggest that couples’ and therapists’ separate worlds of meaning and interpretation result in different appraisals of process and outcome. As Steier (1985) noted, “this understanding is precisely what researchers should be striving for when they study changes that co-occur with family therapy” (p. 32).

Second, it was discussed in the Sells et al. (1994) study that couples and not therapists insisted that reflecting teams should have representatives from both genders. However, the content analysis and further textual examination in the present study revealed that therapists were as concerned as clients as judged by comparable proportional frequency within this category. By comparing numeric data with narrative examples, it was discovered that both therapists and couples reported that having men and women on the team promoted a sense of balance or equality. Having both genders present invoked the feeling of not being intimidated by the opposite sex and promoted different perspectives from a woman’s and a man’s point of view. Moreover, couples reported that men and women “see things differently,” and therefore it was beneficial to have both genders present. Ironically, the Sells et al. (1994) ethnography captured novel aspects of reflecting team practice but might have overlooked
others that were so important that therapists assumed they did not need discussion. After reading this study, several therapists commented that the importance of gender was assumed and did not warrant discussion. The implication is that ethnographic results, like all research findings, should be interpreted with caution and require replication and further analysis before bold claims are made.

Finally, perhaps most intriguing was how couples described their process of being able to sit back and hear what the team was saying to them. In a previous reflecting team study, Prest et al. (1990) conceptualized such a process as a unique “fly on the wall” phenomenon in which someone is at a meta-level to the process. In essence, they were outside the problem and watched someone else talk about that problem without being able to interrupt. This was a unique experience and one that allowed couples to look differently at their concerns.

Statistical and narrative data in this study support the Prest et al. (1990) observations. This sense of being outside the problem was exemplified in the “Spatial Separateness” code where couples reported instances and effects of this boundary between themselves and the team. The process of actually hearing the problem differently was the sequence of steps or stages that couples reported. Although these concepts were sometimes reported as isolated from one another, more often than not a phi coefficient of .78 indicated a close association between the two. Couples repeatedly reported that not being able to object and being forced to go outside the problem allowed them to listen to it differently. Couples offered hints as to the reasons for this occurrence. For some, it was having to hear without interruption the entire discussion of the problem: “At first I was shocked. I didn’t know how to take it, until I stopped being insulted and started listening to what they were saying. And that had all the validity.” For other couples, it gave them time to reflect and see three or four opinions and “objectively dissect each one” after weighing all the options. A close examination of the most frequently used words in context reveals that people “see,” “hear,” or “understand” the problem “differently” because they “can’t” object and they are “outside” the “problem.” A semantic examination of each word in quotes (i.e., KWIC) indicated a close affinity between “Spatial Separateness” and “Process of Hearing” categories in the creation of a different understanding or new reality of the problem.

**Implications: Clinical and Research Integration**

For reflecting team practitioners, effectiveness may depend on the ability of spouses to form multiple perspectives on an impasse. Spatial separateness and the structural process of having spouses listen to a dispassionate discussion of their difficulties provide an experience altogether different from the battle between spouses to ascertain which story is “correct.” The team discussion satisfies a typical request by embattled spouses for an impartial judge (or jury) but does not create the moral and pragmatic difficulties of assuming a referee’s position. Further, the use of reflecting team discussions provides a pause in what may be an impasse between the spouses. For the couple’s impasse to dissolve, such a pause provides a breather in which multiple perspectives are discussed. Because the spouses are not allowed to react immediately to what each might consider to be “wrong,” the enforced pauses allow an opportunity for multiple perspectives to be presented and considered. If nothing else, reflecting team discussions slow down conversations between spouses in which they would ordinarily make repetitive and inflammatory assertions and accusations. Obviously, avoiding such assertions and accusations is crucial to the dissolution of the index problems. One implication for practice is that individuals without teams may experiment with slowing down the interchange between combative spouses and derive benefit by creating therapeutic pauses.
Other implications for reflecting team practice include the following: (a) the index therapist’s use of the reflecting team, (b) gender issues, and (c) couples’ recommended and contraindicated use of the team. Therapists commented frequently that the team helped them “get unstuck” and provided them with a novel therapeutic direction. One therapist stated: “I was aware of how stuck I was feeling with how to move the couple further along. The team helped me get unstuck by giving me a different way to look at the problem.” Other therapists added that team commentary allowed them to take a break to gather their own thoughts or to interrupt the couple’s process before it escalated out of control. Second, both couples and therapists remarked on the importance of having a balance of male and female perspectives on the reflecting team. Spouses observed that each one might have felt “ganged up on” or that “only one point of view would be expressed.” Couples asserted that men and women perceive events differently and that gender representation was important. On reflecting about the process of RT, couples recommended that the team be present within the room and not behind the mirror. When couples could see therapists’ faces, couples found therapists less intimidating. Most couples did not recommend the team for a first session but only after they had begun to trust their own therapist and when they were at an impasse.

Apart from implications for reflecting team practice, we believe that a crucial contribution of the study was in the doing of multimethodological research that integrated research, theory, and practice (Liddle, 1991; Moon & Sprenkle, 1992). Our study attempted to answer the call (Boss, 1992; Moon et al., 1991; Moon & Sprenkle, 1992) for research that is sensitive to clinical exigencies and uses analytic procedures that permit replication and expansion. We believe, as do many others, that family therapy research is enriched by the integration of qualitative and quantitative methods. Ethnographic content analysis (ECA) is unique because it is a blend of qualitative and quantitative methodologies that are grounded in client statements, informed by clinicians’ beliefs, and analyzed by replicable methods.

In essence, we took a methodology that presents information from an etic perspective (researcher’s categories, analyses, etc.) and adapted it to capture an emic perspective (i.e., the worldview of our informants). It would be rash, however, to suggest that the study was purely emic; indeed, the study complemented an etic methodology (content analysis) with data from a previously conducted study that used an emic methodology (the domain analysis reported in Sells et al., 1994). If anything, our goal was to bridge the distinctions between etic/emic and quantitative/qualitative methodologies.

This study was circular rather than linear in nature. We began with inductive inquiry, followed with deductive analysis, started again with more inductive research, and ended with yet more deductive analysis. Such efforts, over a period of years, can result in replicable yet clinically relevant research; our experience suggests that such efforts require persistence and adherence to a research protocol. This study attempted to go beyond dogmatic assertions of belief in any one methodology and to bridge multiple types and levels of inquiry. We believe that multimethodological research models, like ethnographic content analyses, are appropriate for narrative data that tap informants’ underlying meaning worlds. It is clear “that ethnography and quantitative approaches complement each other: one is incomplete without the other” (Werner & Schoepfle, 1987, p. 68). To this end, the story of the informant is not lost but enriched as conformation of previous assertions is revealed and new themes discovered through both deductive and inductive processes.
REFERENCES


qualitative research revisited. *Journal of Marital and Family Therapy*, 17, 187-192.


NOTES

1Within this category, the index therapist describes days in which the reflecting team assisted him or her to resolve an impasse. The term index therapist is adapted from Wynne’s (1988) discussion of how to keep track of clients in family therapy. He argued in favor of index patient for record-keeping versus identified patient. By the same argument, we referred to the lead therapist as the index therapist for record-keeping purposes.

2More information about revising and adding categories is presented later in this paper in the section entitled, “Revising and Implementing the Coding System.” More information about category construction, including terms within each category, and the process used to derive them are available from the authors.

3Strata were based on when the interviews took place; the rationale for using dates of interview as the strata is provided later in the manuscript.