

Field Methods

<http://fmx.sagepub.com>

Intercoder Reliability for Validating Conclusions Drawn from Open-Ended Interview Data

Karen S. Kurasaki

Field Methods 2000; 12; 179

The online version of this article can be found at:
<http://fmx.sagepub.com/cgi/content/abstract/12/3/179>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Field Methods* can be found at:

Email Alerts: <http://fmx.sagepub.com/cgi/alerts>

Subscriptions: <http://fmx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Intercoder Reliability for Validating Conclusions Drawn from Open-Ended Interview Data

KAREN S. KURASAKI
University of California, Davis

Intercoder reliability is a measure of agreement among multiple coders for how they apply codes to text data. Intercoder reliability can be used as a proxy for the validity of constructs that emerge from the data. Popular methods for establishing intercoder reliability involve presenting predetermined text segments to coders. Using this approach, researchers run the risk of altering meanings by lifting text from its original context, or making interpretations about the length of codable text. This article describes a set of procedures that was used to develop and assess intercoder reliability with free-flowing text data, in which the coders themselves determined the length of codable text segments. Content analysis of open-ended interview data collected from twenty third-generation Japanese American men and women generated an intercoder reliability of more than .80 for fifteen of the seventeen themes, an average agreement of .90 across all themes, and consistency among the coders in how they segmented coded text. The findings suggest that these procedures may be useful for validating the conclusions drawn from other qualitative studies using text data.

Qualitative researchers have long been criticized for providing less than adequate evidence for the conclusions they draw. Researchers who use text data are vulnerable to committing systematic and nonsystematic errors throughout the complex tasks of developing and applying codes to the data. To avoid falling victim to such errors, many researchers (e.g., Bernard 1995: 342–43) have suggested using multiple, well-trained coders and establishing high intercoder reliability.

Intercoder reliability is a measure of agreement between multiple coders about how they apply codes to the data. Agreement can be used to measure the reliability of the coders as instruments to identify and mark themes in a text, or as a proxy for the validity of the constructs that emerge from the data (Ryan 1999). In the latter capacity, agreement demonstrates that the themes

This article was prepared with the partial support of the National Research Center on Asian American Mental Health (National Institutes of Mental Health grant no. R01 MH44331). The author wishes to acknowledge Gery Ryan, Roderick J. Watts, and Doris Chang.



Field Methods, Vol. 12, No. 3, August 2000 179–194
© 2000 Sage Publications, Inc.

are shared constructs and not simply a figment of the investigator's imagination.

Intercoder reliability is calculated by examining the degree to which coders agree across a fixed set of units. In classic content analysis, these units are often predetermined. They may be short-answer responses to open-ended questions or linguistic units such as pages or paragraphs taken from written texts such as newspaper columns (for additional types of coding units, see Krippendorff 1980). Researchers who want to calculate intercoder reliability for free-flowing texts face a fundamental problem: What are the units to be coded?

One solution is to identify "idea units" based on where a single idea starts and ends. This approach aids the researcher in reducing the data into more manageable chunks that, in turn, can be presented to coders for classical content classification or coding (Carey, Morgan, and Oxtoby 1996). Unfortunately, respondents do not speak in text units. Typically, narrative data are conversational and quite choppy. Such interviews may revolve around one main topic at a time, but may also include frequent tangents, digressions, backtracks, and overlaps. The question then becomes: How does the investigator lift segments of text from their original context and still maintain the integrity of its intended meaning? Bounding the data into more manageable text units while facilitating the task of conducting intercoder reliability risks pulling too much out of its original context and requires the researcher to impose too much interpretation before the data reach the coders.

This article offers an alternative solution for calculating intercoder reliability on free-flowing texts. Instead of having assistants code predetermined text units, coders were allowed to mark the free-flowing text as they felt was appropriate, and intercoder reliability was then calculated by checking the agreement between coders at random points in the text.

EXAMPLE

I demonstrate the reliability of this method for assessing intercoder reliability with data from a psychological study on the personal meanings associated with being Japanese American. The goal of the study was to build theory for understanding the meanings people apply to their life experiences as *sansai*, or third-generation Japanese Americans. Most research on Japanese Americans has focused on the acculturation of earlier generations to the dominant host culture (Meredith 1967; Masuda, Matsumoto, and Meredith 1970; Okano and Spilka 1971; Pierce, Clark, and Kaufman 1978–79; Newton et al. 1988). The life experiences of Japanese Americans have largely

been reduced to issues revolving around English-language acquisition and adjustment to Western values, social norms, and customs. Other theoretical frameworks for understanding the Japanese American experience have been neglected.

Two theoretical frameworks in particular—racial identity and ethnic identity—warrant attention. Racial identity has been described as both a self-concept (Thomas 1971; Cross 1978; Parham 1989) and a worldview (Helms 1993) that is formed on the basis of being a racial minority. An examination of social and political events in this country demonstrates a history of systematic discrimination against Japanese Americans, suggesting the potential relevance of racial identity among Japanese Americans.

Unlike racial identity, ethnic identity is neither a self-concept nor a worldview, but a sense of belonging (Tajfel 1981; Phinney 1990). Ethnic identity represents a sense of kinship with other people who share similar ethnic origins and cultural heritage. Most Japanese Americans today descend from immigrant ancestors who came to the United States from Japan in the late 1800s and early 1900s. The extent to which Japanese Americans retain traditional Japanese customs and values today may provide a common understanding that serves as the basis of kinship and bonding among Japanese Americans. Because neither of these theoretical constructs—racial identity nor ethnic identity—has been explored among Japanese Americans, I used an exploratory approach in the work reported in this article.

DATA COLLECTION METHOD

Sample

Twenty third-generation Japanese American men and women were recruited for the study in the greater Los Angeles area using snowball sampling (Patton 1990). That is, early recruits led me to the later recruits. Not every person who was referred to the study was ultimately selected for participation. Instead, I chose participants based on three criteria: (1) Participants had to be Japanese American on both their maternal and paternal sides, but had to be third generation based on either maternal or paternal history; (2) participants had to have at least one parent who was interned in the concentration camps run by the U.S. government during World War II; and (3) participants had to be between the ages of thirty and fifty.

Twenty-six eligible persons were either self-selected, recruited by me, or referred by others to the study. Four of these twenty-six refused participation and two dropped out after committing to the interview, resulting in twenty

completed cases (attrition rate 24%). Of these, three were self-selected, eight were directly recruited, and nine were referred by others.

Sample Characteristics

The sample consisted of ten women and ten men who ranged in age from thirty-two to forty-nine years at the time of the interview. All participants were educated beyond high school, although more than half came from families where the highest educational level attained by at least one parent was less than a college bachelor's degree. The majority ($n = 15$) of the participants were currently married or were living with a partner. Of the remaining five participants, four had never been married and one had been widowed. Nearly all ($n = 18$) had minimal or no proficiency in the Japanese language. Similarly, most individuals reported minimal ($n = 11$) to no ($n = 6$) involvement in Japanese American community activities, whereas three reported participating more extensively in the Japanese American community.

Interviews

Semistructured interviews were conducted with each informant in English. The interview format was open ended, facilitating a free flow of ideas from the respondent and generating information-rich data. I adopted a clinical method of interviewing (Dapkus 1985) for this project to understand the meanings Japanese Americans have created from their life experiences. In this method, the interviewer asks the respondent to elaborate on a statement if it seems ambiguous, until both the interviewer and the respondent feel mutually satisfied with the clarity of a concept.

All the interviewers involved in the project used an interview protocol to ensure that all the research questions in my design would be covered. The research questions were based on existing theory and measures (Phinney 1990; Nagata 1993), on my preliminary investigation (Kurasaki 1994), and on data from a pilot study for this project. The interview covered three main areas: (1) how people defined themselves, ethnically; (2) what the components were of this ethnic identity; and (3) how this ethnic identity developed in the course of people's lives. To get at this last component, the interview focused on (a) transitions in ethnic identification across the life span and (b) situational influences on the identification process. On average, interviews lasted about ninety minutes.

The analytic techniques presented here were developed to address the second area: defining identity components or themes. This portion of the interview began with the grand-tour question: "What sorts of things remind you

that you're Japanese American?" The question was intended to elicit responses that would lead to definitions of identity components or themes.

There were three interviewers on this project, including myself. Although the interviewers were somewhat younger than the respondents overall, interviewers were matched to respondents by gender and generational status. All the interviewers received extensive training in establishing rapport, the use of silence and echoes, probing, and other interviewing techniques described by Bernard (1995). The interviews were conducted in a location requested by respondents themselves. Most were conducted in the respondent's home or place of work. All interviews were audio recorded.

The interviews were transcribed verbatim, following a standardized set of typing procedures. There were two levels of reviewing and editing. First, research assistants (noninterviewers on the project) did a thorough review of each transcript, matching it against the tape. Second, I performed a final review and edit of each transcript.

DATA ANALYSIS METHOD

A team of researchers—myself and the two research assistant/interviewers—analyzed the verbatim transcripts, identifying themes related to ethnic identity. As shown in Figure 1, this analysis occurred in three phases. In the first stage, the researchers identified potential themes and developed a formal codebook. In the second stage, the team used a sample of the text to establish intercoder reliability for each of the themes in the codebook. Finally, once the researchers were confident that they had established acceptable levels of intercoder reliability, they proceeded to the final stage and systematically applied the codebook to the entire corpus of text. Each of these steps is discussed below.

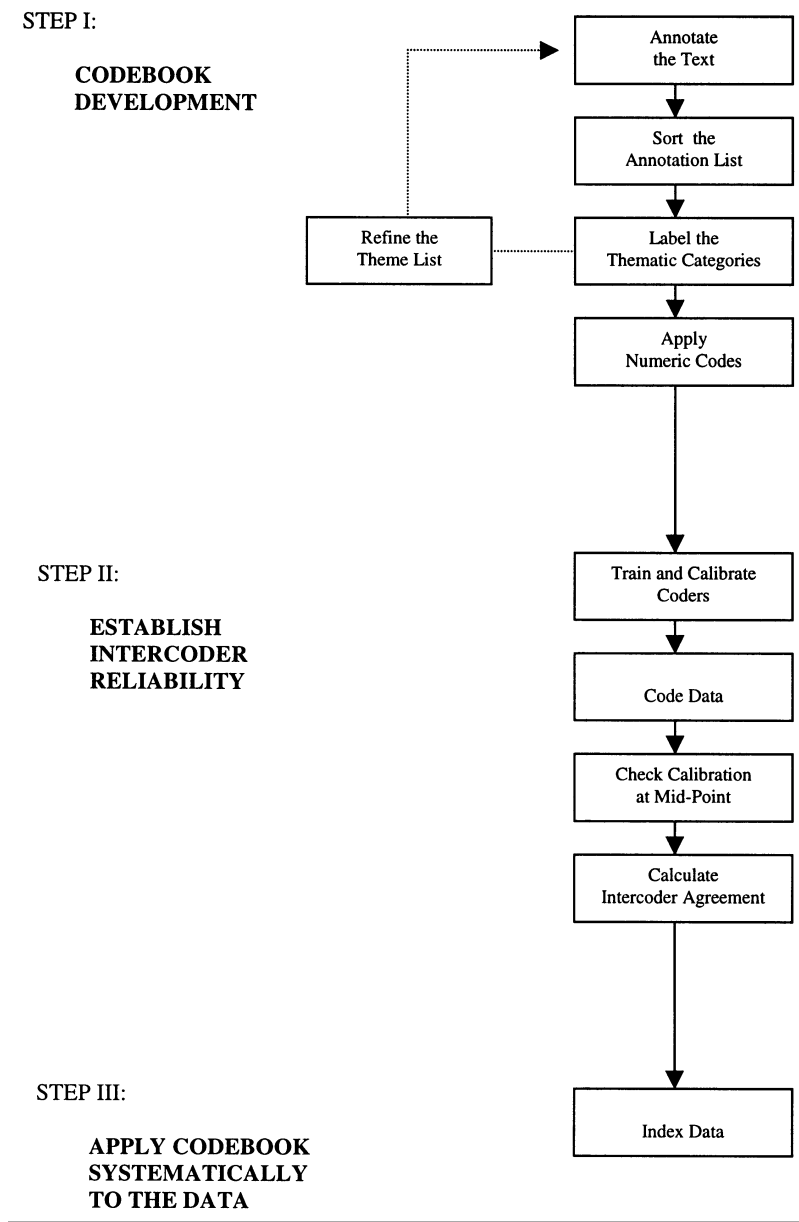
Step 1: Codebook Development

The research team members used a grounded theory approach (Strauss and Corbin 1990) to identify potential themes and to develop a formal codebook. The following section describes the steps that were followed to implement grounded theory systematically.

Annotating the Text

The research team examined the first five transcripts together and performed a step called annotation. Annotations are notes written in the margins

FIGURE I
Thematic Analysis of Free-Flowing Text Data



with regard to the interview's contents. Generally, the annotations were brief, one- to ten-word notes that summarized the main points expressed. For example, for the following excerpt, a typical annotation would have been "traditional Japanese customs from mom":

Oh, silly things like my mom had told me whenever someone brings you something, you never return the bowl empty, you know, you always put something back in to say thank you.

As the research team proceeded with the annotations, the same themes began to emerge repeatedly. By the time we had gone through five of the transcripts, agreement was achieved across all three researchers on how to annotate the themes that were emerging. At this point, I began to annotate independently the remaining fifteen transcripts.

Sorting the Annotation List

Next, I compiled a verbatim list of all the annotations, which I called the annotation list. I extracted very short verbatim quotes from the raw data (i.e., one sentence maximum) that served as examples for each annotation in the annotation list. Then, all three members of the research team sorted the annotations in this list into similar categories and subcategories. Working together, we were able to eliminate redundancies and create a preliminary hierarchy of thematic categories. As an outcome of this consolidation step, twenty-nine thematic categories were identified.

Labeling the Thematic Categories

For each of the twenty-nine thematic categories, the research team developed descriptive labels to represent the intended meaning of each category. These short (i.e., one to five words in length), descriptive labels, which I called the theme list, served as the foundation for the codebook that would be operationalized in step 2.

Refining the Theme List

Using the theme list as a guide, I indexed all transcripts on a computer file. Once all the data were indexed, I did searches using the different themes as the search criteria.¹ This generated separate reports, one for each of the twenty-nine themes. Each report contained all the indexed data (i.e., verbatim excerpts that ranged from one or two words to several pages in length) that were retrieved in the search for its theme.

TABLE I
Theme List and Final Codebook

| <i>Code</i> | <i>Theme List</i> | <i>Final Codebook</i> |
|-------------|--|--|
| 1.1 | Sense of having a Japanese heritage | Sense of having a Japanese heritage |
| 1.2 | Sense of having a Japanese American social history | Sense of having a Japanese American social history |
| 2.1 | Japanese American values and attitudes | Japanese American values and attitudes |
| 2.2 | Practice of Japanese customs | Practice of Japanese customs |
| 2.3 | Japanese ways of doing things | Japanese ways of doing things |
| 2.4 | Japanese American interpersonal or communication styles | Japanese American interpersonal or communication styles |
| 2.5 | Japanese-language proficiency | Japanese-language proficiency |
| 3.1 | Integration or multicultural competence | Integration or multicultural competence |
| 3.2 | Multicultural conflict or confusion | Multicultural conflict or confusion |
| 4.1 | Sense of a global ethnic or racial community | Sense of a global ethnic or racial community |
| 4.2 | Sense of interpersonal connectedness with same ethnicity or race others | Sense of interpersonal connectedness with same ethnicity or race others |
| 4.3 | Sense of intellectual connectedness with other ethnic or racial minorities | Sense of intellectual connectedness with other ethnic or racial minorities |
| 4.4 | Searching for a sense of community | Deleted |
| 4.5 | Sense of alienation from ascribed ethnic or racial group | Sense of alienation from ascribed ethnic or racial group |
| 5.1 | Sense of comfort with one's ethnic or racial self | Sense of comfort with one's ethnic or racial self |
| 5.2 | Searching for a sense of comfort with one's ethnic or racial self | Searching for a sense of comfort with one's ethnic or racial self |
| 6.1 | Social consciousness | Orientation toward promoting racial tolerance |
| 6.2 | Sense of oppression | Sense of oppression |

NOTE: Modifications that were made to the theme list are indicated in bold in the final codebook.

This step of generating an individual report for each theme was intended to aid in the examination of the taxonomy for possible redundancies missed during the earlier sorting step. Indeed, by repeating the annotation step on these twenty-nine individual reports, the research team observed several redundancies and, after further sorting, selected eighteen themes for further consideration. Table 1 lists those themes.

Applying Numeric Codes

A decimal numbering system was applied so that the numerical digits could represent different levels of the themes. For instance, both of the themes that were assigned a 6.x numeric code pertain to one's worldview, that is, the filter through which one observes and interprets events and experiences. The digit after the decimal further distinguishes between two different types of worldviews, although in no particular order. The code 6.1 was applied to the theme that represents seeing one's experiences and position in life within the context of racial oppression. The code 6.2 was applied to the theme that represents a commitment to promoting interracial understanding. The unique numeric identifier or code assigned to an individual theme is shown in Table 1.

Final Codebook Refinement

Further refinement of the codebook occurred while training four research assistants to be coders. During the training session, coders were familiarized with the definitions and use of the themes, and the coding procedures. Two amendments were made to the codebook at the coders' request. They unanimously opted to eliminate category 4.4 ("searching for a sense of community") due to its redundancy with 4.1 ("sense of a global ethnic or racial community"). They also reworded category 6.1 from "social consciousness" to "orientation toward promoting racial tolerance," which they agreed better reflected its intended meaning. As a result of these amendments, the final codebook consisted of seventeen categories and their respective numeric codes (see Table 1).

Step 2: Establishing Inter-coder Reliability

Establishing inter-coder reliability comprised several procedures. These procedures are described in the following section.

Training Procedures

One of the original four coders dropped out early in the training, leaving three research assistants who were fully trained to perform the coding tasks. The remaining three coders included two persons who had interviewed participants and who had participated in developing the codebook. The third coder had performed the transcript checking and editing tasks. Thus, all coders were very familiar with the data. To test their understanding of each code, all three coders read the same portions of five randomly selected interviews

and marked the text for all seventeen themes. The marking behaviors of the coders were compared to calculate measures of intercoder reliability. The three coders were calibrated to an agreement level of .70 on the data presented to them during the training phase.

Coding Procedures

Each coder was given the same set of twenty transcript pages. He or she was then instructed to independently delineate text units and assign each text unit a numeric code or codes. A text unit was defined as a segment of conversation that represented a single message (McFadden, Seidman, and Rappaport 1992), a distinguishing feature (Pennartz 1986), or a change of subject (Dapkus 1985).

A random cluster sample was used to select which transcript pages were to be coded. I randomly selected five (25%) of the transcripts and then randomly selected four pages from each of the five randomly selected transcripts. Averaging across all twenty transcripts, the average number of pages was sixteen. Thus, four pages represented 25% of an average transcript.

Agreement Calculation

To calculate agreement, I randomly selected ten different lines per coded page. For each randomly selected line, or any line within five lines above or below the randomly selected line, I examined the marking behavior of the three coders for agreement. The reason for including five lines above and below the actual line was to accommodate the variation that was expected to occur in how each of the three coders bounded the data. Because of the conversational nature of the interviews, it was common to find peripheral text surrounding more substantive, codable text. Consider, for example, the following:

I've been thinking a lot about this. I mean, everybody's thinking about this nowadays, right? And, and I think for an organization like the church it's a very central question. And, and particularly now with the presence of inter-, more intermarriage families, and more biracial folks, and we used to call them *hapas*, but more biracial people. I mean, it's a real important issue to deal with. And, and I go back and forth thinking about this all the time. I think, let me just in a nutshell, if I could articulate this, there is such a thing as a Japanese American identity. That, that is my starting thesis. However, I think the definition needs to be expanded from being an identity based on race and blood lineage, it needs to be expanded from that to a broader understanding of an identity that may or may not include race or blood. But, an identity that is able to embrace the experience and the story of, of Japanese Americans.

TABLE 2
Intercoder Agreement for the Seventeen Themes

| <i>Agreement</i> | <i>Code</i> | <i>Theme</i> |
|------------------|-------------|--|
| .68 | 1.1 | Sense of having a Japanese heritage |
| .68 | 1.2 | Sense of having a Japanese American social history |
| .84 | 2.1 | Japanese American values and attitudes |
| .95 | 2.2 | Practice of Japanese customs |
| .99 | 2.3 | Japanese ways of doing things |
| .98 | 2.4 | Japanese American interpersonal or communication styles |
| 1.00 | 2.5 | Japanese-language proficiency |
| 1.00 | 3.1 | Integration or multicultural competence |
| .85 | 3.2 | Multicultural conflict or confusion |
| .88 | 4.1 | Sense of a global ethnic or racial community |
| .92 | 4.2 | Sense of interpersonal connectedness with same ethnicity or race others |
| .97 | 4.3 | Sense of intellectual connectedness with other ethnic or racial minorities |
| — | 4.4 | Deleted |
| .95 | 4.5 | Sense of alienation from ascribed ethnic or racial group |
| .97 | 5.1 | Sense of comfort with one's ethnic or racial self |
| .86 | 5.2 | Searching for a sense of comfort with one's ethnic or racial self |
| .82 | 6.1 | Orientation toward promoting racial tolerance |
| .89 | 6.2 | Sense of oppression |

The first five lines of text ultimately lead to the general theme of a Japanese American identity, although they do not stand alone as codable text. In coding this excerpt, then, a more inclusive coder might bound the entire text as one text unit. In contrast, a less inclusive coder might begin coding the text from “there is such a thing as a Japanese American identity.” Thus, variation in how coders segmented the text was anticipated.

A 1 in the agreement matrix (see Figure 2) indicated that a code had been applied to the randomly selected line (or any line within five lines above or below the randomly selected line), and a 0 indicated that the code had not been applied. Agreement across the three coders for each of the themes was calculated by using a ratio of agreements to disagreements. An overall agreement across all the themes was calculated by averaging the agreements obtained for each theme.

Calibration Check at Midpoint

As an added precaution to ensure that the coders' comprehension of the codes remained consistent throughout the coding process, I calculated an

FIGURE 2
Agreement Matrix

| line | coder | 1.1 | 1.2 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | 3.1 | 3.2 | 4.1 | 4.2 | 4.3 | 4.5 | 5.1 | 5.2 | 6.1 | 6.2 |
|------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | A | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | |
| | C | | | | | | | | | | | | | | | | | |
| | A | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | |
| | C | | | | | | | | | | | | | | | | | |
| line | coder | 1.1 | 1.2 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | 3.1 | 3.2 | 4.1 | 4.2 | 4.3 | 4.5 | 5.1 | 5.2 | 6.1 | 6.2 |
| | A | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | |
| | C | | | | | | | | | | | | | | | | | |
| | A | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | |
| | C | | | | | | | | | | | | | | | | | |
| line | coder | 1.1 | 1.2 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | 3.1 | 3.2 | 4.1 | 4.2 | 4.3 | 4.5 | 5.1 | 5.2 | 6.1 | 6.2 |
| | A | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | |
| | C | | | | | | | | | | | | | | | | | |
| | A | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | |
| | C | | | | | | | | | | | | | | | | | |

overall agreement across all coders and themes after half of the data had been coded. Agreement on the data coded up to that midpoint was .94, which was determined to be high enough for coders to finish coding the remaining data without any intervention.

Intercoder Agreement for the Seventeen Themes

The intercoder agreement reported in this section is based on the marking behavior of all three coders and on all twenty completed transcript pages. Intercoder agreement for each of the seventeen thematic categories was calculated, as well as an overall average agreement across all seventeen themes. Intercoder agreement on the seventeen themes ranged from .68 to 1.00, for an average of .90. These results are summarized in Table 2.

Examining Text Segmentation Patterns

MacQueen et al. (1998) lists two types of coding inconsistencies: (1) inconsistencies in the application of codes within segments and (2) inconsistencies in defining the beginning and end of segments. Intercoder agreement described earlier in this article was calculated to examine the former inconsistency. To examine the latter inconsistency, that is, inconsistency in segmenting text, I studied the text segmentation patterns of the three coders.

The procedures for obtaining intercoder agreement in this study involved allowing the coders to mark free-flowing text, a technique that has not been previously tested. To more fully understand the utility of this alternative technique, I examined whether the text segmentation patterns of the three coders affected the obtained intercoder reliability.

Inconsistency in text segmentation was examined by calculating two scores. First, the total number of text segments across all twenty pages was summed for each coder. Second, the mean or average number of lines per segmented text was calculated for each coder and compared in an analysis of variance (ANOVA). The total number of text segments for coders A, B, and C were 32, 30, and 29, respectively. The mean number of lines per segmented text were 17.94, 18.17, and 23.07, respectively, for coders A, B, and C. An ANOVA indicated no significant difference between the three coders on the mean number of codes per segmented text. These findings, which are presented in Table 3, suggest that the variation in text segmentation found among the three coders did not significantly affect the obtained intercoder reliability.

Step Three: Applying the Codebook Systematically to the Data

Recall that intercoder reliability was obtained using a portion of all the data. Intercoder reliability was established in this manner to confirm that the

TABLE 3
Text Segmentation Patterns

| | <i>Coder A</i> | <i>Coder B</i> | <i>Coder C</i> |
|----------------------------------|----------------|----------------|----------------|
| Text segmentation | | | |
| Total number of segments | 32.00 | 30.00 | 29.00 |
| Mean number of lines per segment | 17.94 | 18.17 | 23.07 |

$p < .05$.

codebook comprised themes that were shared constructs. Having obtained an acceptable level of intercoder agreement, I indexed the entire data set using the agreed-upon seventeen themes. There were two reasons for this. (1) I wanted to be able to perform searches of all the data by theme and to count the number of hits. The frequency counts served as an indication of how prominent each thematic category was among the sample. (2) I wanted to be able to retrieve quotes from the data for presentations and publications—quotes that represented each of the seventeen thematic categories. The findings from this third step are not reported here.

DISCUSSION

To overcome criticism that qualitative analysis is overly subjective, several researchers have recently asserted that scientists need to do more than report explicitly on the procedures they employ for analyzing their text data (MacQueen et al. 1998). They need also to implement systematic steps such as intercoder agreement (Carey et al. 1996) to validate their conclusions. Systematic, qualitative analysis of text data involves assigning unique labels or codes to text passages. Intercoder agreement is a step that helps ensure that coding is reliable or, in other words, replicable. Intercoder agreement is demonstrated when multiple coders can independently replicate each other's work of assigning codes to text data. High intercoder agreement or intercoder reliability strengthens confidence among the scientific community that the theoretical conclusions drawn from text data analysis are valid.

One of the more popular approaches for establishing intercoder reliability is one in which the researcher presents predetermined text units to the coders. That is, the researcher first segments the data into manageable chunks, producing finite units of text that can then be presented to the coders for the task of coding. One of the problems with the researcher segmenting text in this manner is that the researcher is then imposing considerable interpretation on

the data before the data reach the coders. Another concern with this approach is that of lifting text from its original context and thereby losing or altering its meaning in the process.

This article describes a set of procedures for developing and assessing intercoder reliability without predetermining the length of coded text segments. Applying these procedures to the analysis of interview transcripts, researchers were able to generate intercoder reliability of more than .80 for fifteen of seventeen coding themes, an average agreement across all seventeen themes of .90, and consistency between three coders in how the coded text was segmented. These findings suggest that the procedures described here can be very useful in assessing and increasing the reliability of coded text data in other studies. Implementing such procedures can strengthen the validity of conclusions based on coded text data and respond effectively to ongoing criticisms about the apparently subjective nature of qualitative methods.

NOTE

A Word macro designed especially for the purposes of indexing and searching text data was developed by Gery Ryan, Department of Anthropology, University of Missouri.

REFERENCES

- Bernard, H. R. 1995. *Research methods in anthropology: Qualitative and quantitative approaches*. 2d ed. Walnut Creek, CA: AltaMira.
- Carey, J. W., M. Morgan, and M. J. Oxtoby. 1996. Intercoder agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *Cultural Anthropology Methods* 8 (3): 1–5.
- Cross, W. 1978. The Thomas and Cross models of psychological nigrescence: A literature review. *Journal of Black Psychology* 4:13–31.
- Dapkus, M. A. 1985. A thematic analysis of experience of time. *Journal of Personality and Social Psychology* 49:408–19.
- Helms, J. E. 1993. *Black and white racial identity: Theory, research, and practice*. Westport, CT: Praeger.
- Krippendorff, K. 1980. *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.
- Kurasaki, K. S. 1994. Japanese American ethnic identity: Psychological and social dimensions. *The Community Psychologist* 27:56–57.
- MacQueen, K. M., E. McLellan, K. Kay, and B. Milstein. 1998. Codebook development for team-based qualitative analysis. *Cultural Anthropology Methods* 10 (2): 31–36.
- Masuda, M., G. H. Matsumoto, and G. M. Meredith. 1970. Ethnic identity in three generations of Japanese Americans. *Journal of Social Psychology* 81:199–207.

- McFadden, L., E. Seidman, and J. Rappaport. 1992. A comparison of espoused theories of self- and mutual help: Implications for mental health professionals. *Professional Psychology: Research and Practice* 23:515–20.
- Meredith, G. M. 1967. Ethnic identity scale: A study in transgenerational communication patterns. *Pacific Speech Quarterly* 2:57–67.
- Nagata, D. K. 1993. *Legacy of injustice: Exploring the cross-generational impact of the Japanese American internment*. New York: Plenum Press.
- Newton, B. J., E. B. Buck, D. T. Kunimura, C. P. Colfer, and D. Scholsberg. 1988. Ethnic identity among Japanese Americans in Hawaii. *International Journal of Intercultural Relations* 12:305–15.
- Okano, Y., and B. Spilka. 1971. Ethnic identity, alienation and achievement orientation in Japanese-American families. *Journal of Cross-Cultural Psychology* 2:273–82.
- Parham, T. A. 1989. Nigrescence: The transformation of black consciousness across the life cycle. In *Black adult development and aging*, edited by R. L. Jones, 151–66. Berkeley, CA: Cobb & Henry.
- Patton, M. Q. 1990. *Qualitative evaluation and research methods*. 2d ed. Newbury Park, CA: Sage.
- Pennartz, P.J.J. 1986. Atmosphere at home: A qualitative approach. *Journal of Environmental Psychology* 6:135–53.
- Phinney, J. S. 1990. Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin* 108:499–514.
- Pierce, R. C., M. Clark, and S. Kaufman. 1978–79. Generation and ethnic identity: A typological analysis. *International Journal of Aging and Human Development* 9:19–29.
- Ryan, G. W. 1999. Measuring the typicality of text: Using multiple coders for more than just reliability and validity checks. *Human Organization* 58:313–22.
- Strauss, A., and J. Corbin. 1990. *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Tajfel, H. 1981. *Human groups and social categories: Studies in social psychology*. New York: Cambridge University Press.
- Thomas, C. W. 1971. *Boys no more: A black psychologist's view of community*. Beverly Hills, CA: Glencoe.

KAREN S. KURASAKI is research faculty in the Department of Psychology at the University of California, Davis. Her research interests include ethnic and racial identity development, mental health systems of care, and community-based prevention interventions to strengthen youth and families. Her recent publications include "Racial Identity and Treatment Issues Among Asian and Pacific Islander Americans," a chapter that appears in Asian and Pacific Islander Americans: Issues and Concerns for Counseling and Psychotherapy (Nova Science Publishers, Inc., 1999), and "Ethnic Minority Intervention and Treatment Research," which appears in Psychological Intervention and Cultural Diversity, Second Edition (Allyn and Bacon, 2000).