Measuring Cultural Consonance: Examples with Special Reference to Measurement Theory in Anthropology

WILLIAM W. DRESSLER
University of Alabama
CAMILA D. BORGES
University of São Paulo
MAURO C. BALIEIRO
Paulista University
JOSÉ ERNESTO DOS SANTOS
University of São Paulo

A valid and reliable anthropological measurement must be culturally appropriate for a particular social setting. Justifying the appropriateness of a measurement often depends on the skill of the researcher in describing the ethnographic setting. This has resulted in valuable research, but it is difficult to systematize and lacks transparency. Here the authors present a measurement model for anthropology that links structured ethnographic methods—cultural domain analysis and cultural consensus analysis—to the assessment of individual behavior and personal beliefs. These procedures are illustrated with the concept of cultural consonance, or the degree to which an individual approximates in his or her own behavior or belief the shared cultural model in some domain. The concrete steps taken to develop measures of cultural consonance in four domains (lifestyle, social support, family life, and national characteristics) are described, and the reliability and validity of these measures are evaluated. This describes a measurement model for anthropology.

Keywords: cultural consonance; cultural consensus analysis; cultural models; measurement theory

This article describes the measurement of individual beliefs and behaviors relative to a particular cultural context. Formal hypothesis testing in anthro-
pology requires the collection of numeric data over a range of sampled entities. Measurement is the principled assignment of numbers to sampled entities. A measurement model is a set of rules for assigning numbers to sampled entities to represent their relative positions on some attribute (Bernard 1994:24–25). In this article, we present a specific set of measures of “cultural consonance.” This illustrates a measurement model useful in anthropology (Dressler 1996; Handwerker 2002).

MEASUREMENT IN ANTHROPOLOGY AND OTHER SOCIAL SCIENCES

Of the social sciences, psychology is most explicit in measurement theory. Measurement models in psychology include both classic test theory (Guilford 1954; Nunnally 1978) and item response theory (Embretson and Reise 2000; DeVellis 2003). Anthropology borrows heavily from psychometric theory, especially classic test theory (Pelto and Pelto 1978:33–34; Bernard 1994:38–43; Handwerker 2001:188–89). This provides the framework for this article. The issue in psychometric theory is how to array individuals along a continuum that represents varying levels of some attribute. Psychometric theory provides procedures to evaluate both the reliability and the validity of such an array.

Measures of attributes are often objected to by anthropologists in specific contexts, even though these measures meet requirements of reliability and at least face validity. Why? Typically, the notion of “meaning” is invoked. Take, for example, a measure of psychological dysfunction that includes a question referring to hearing voices. In a society in which visits by unseen entities are regarded as part of normal life (e.g., among spirit mediums in candomblé in Brazil), this item may not measure dysfunction since the experience is culturally regarded as normal. It signifies something different (R. Cohen and Naroll 1973:16; Poortinga 1989).

In psychometric theory, this is a question of sampling the universe of items intended to measure some attribute. In practice, scale items are generated on the basis of theoretically informed expectations that responses to the items reflect the attribute. The items are assumed to be a random sample from a universe of items that can serve as indicators of an attribute (DeVellis 2003:64–65). The example given above shows that anthropologists balance two concerns. One of these, of course, is theory. The other concern is local meaning. What, in terms of local meaning, is an appropriate indicator of the
variable in question? Dressler (1995) recommends an “ethnographic critique of theory,” or examination of how a theory is instantiated within a specific cultural context. This manifestation begins with locally meaningful indicators of relevant variables.

Traditionally, anthropologists have approached this question with the conventional tools of ethnography. Janes’s (1990) study of social stress and blood pressure among Samoan migrants to Northern California is a good example. Janes’s theoretical orientation led him to examine both how Samoans aspire to higher social status in their community and the limitations placed on achieving that status (this incongruence hypothesized to be stressful). One avenue of status aspiration is receiving a title as matai, or political leader. To make his case, Janes had to demonstrate the importance of the matai status in ethnohistorical materials for Samoa and then, via key informant interviews and participant observation, show that this status position had been transplanted to the American context. Then, he chose to use achievement of matai status as one measure of status aspiration.

Assessing the adequacy of this measure depends on the reader of the research being convinced that this measure “makes sense” in Samoan migrant culture. In part, it becomes a rhetorical issue to artfully embed the measure in cultural context (which Janes does admirably). Then, the measure is correlated with other indicators and with antecedent or outcome variables with which it is hypothesized to be associated (Janes’s measure performs very well in this respect).

There is, however, a lack of transparency in the measurement process. It might be difficult for another ethnographer, lacking Janes’s skill, to replicate his procedures. There are also untested assumptions. First, the degree of knowledge sharing is not tested. The approach to measurement illustrated by Janes’s (1990) study assumes that the relevant items regarding status aspirations are widely understood to be such by members of the Samoan community. Second, it is assumed that the indicators of status aspiration are equal in their cultural importance since each received an equivalent weight in measurement. (Using Janes’s study as an example here should not be construed as a lack of admiration for it. Many other works could be subjected to the same critique with respect to measurement issues.)

The value of anthropological research would be enhanced by systematic measurement procedures; however, considerable effort must be taken not to lose the very thing that makes hypothesis testing in anthropological research most useful, which is the sensitivity to local meaning and context in measurement.
A NEW ANTHROPOLOGICAL MEASUREMENT MODEL

The approach illustrated here was suggested by Dressler (1996) and elaborated by Handwerker (2002). This approach uses recent innovations in anthropological theory and methods to formally test for shared understanding in a cultural domain. This information is then used to construct measures of individual behavior within that domain.

Bourdieu’s (1984) Euclidean model for culture and social structure is a useful framework for this measurement model (see also Crossley 2001). Bourdieu uses the term *cultural space* to describe the culturally constructed world as it is understood by its inhabitants. In one sense, conventional psychometric approaches to evaluating measurement examine only the adequacy of a measure within a cultural space. These approaches do not determine if a measure is appropriate for a given cultural space. Handwerker (2002) suggests a means of first examining the structure of that cultural space.

To begin, a precise explication of culture as shared meaning is needed. This perspective has been elaborated in cognitive anthropology over the past fifty years (Holland and Quinn 1987; D’Andrade 1995; Shore 1996). Although agreement is not complete, a working theory can be delineated. Culture is not regarded as an integrated whole but as a set of cultural models for various cultural domains. These models are skeletal outlines of the elements of the domain and basic processes within the domain but leave many variables to be specified within particular instances. Individual models have two components: One is a function of individual biography; the other is a function of what the individual learns about that domain as a member of society and is a cultural model because it is shared with other members of society (Shore 1996:49).

The notion of sharing or consensus is essential, as has been recognized for more than a century in the social sciences (Tylor 1871; Berger and Luckman 1967). Many (although probably not all) cultural models define things in the world in an essentially arbitrary way. What gives these arbitrary definitions causal force is that people agree that this is, indeed, the way things are (D’Andrade 1984).

Understanding the importance of consensus is essential; defining it empirically is another matter. People will agree on the nature of cultural things to a degree, leaving room for some models to be highly contested, while others are accepted with little dispute. Romney, Weller, and Batchelder (1986) introduced the cultural consensus model, which formally quantifies consensus. (A formal derivation of the cultural consensus model can be found in Batchelder and Romney [1988]; and Romney and Batchelder [1999] provide additional references to the formal elements of the model.)
Cultural consensus theory assumes a fixed knowledge base for questions that are meaningful in a particular social context and that individuals are differentially able to access that fixed knowledge base. Working from the pattern of agreement among key informants, the cultural consensus model determines the degree of sharing in a domain. The degree of consensus in a domain enables the analyst to infer within certain confidence limits that these informants are, or are not, operating from a shared cultural model. In addition, the cultural consensus model can operationalize the degree to which individuals share in the overall consensus. This is the concept of cultural competence, which is the correlation between an individual's understanding of the domain and the consensus understanding of the domain. It should be emphasized here that the assessment of cultural consensus implies the assessment of intracultural diversity; the range and standard deviation of the cultural competence coefficients are measures of diversity (Jaskyte and Dressler 2004).

Finally, the cultural consensus model enables the analyst to estimate the "culturally best" set of responses within a particular domain. The responses are estimated giving higher weight to informants who have higher cultural competence (Romney, Weller, and Batchelder 1986). This characteristic of the model is particularly important, not only because it is a culturally unbiased estimate but also because with those estimates, the elusive aggregate quality of culture can sensibly be grasped. The culturally unbiased responses estimated from the model are not an average but take into account how meaning is distributed among informants.

The results of a cultural consensus analysis represent how respondents act on their knowledge of the cultural model and enable the analyst to make inferences about it. But the results of cultural consensus analysis are not the model hypothesized to exist in individuals’ minds; these results are, rather, an outcome of individuals working from that model.

The cultural consensus model represents a well-developed measurement theory for assessing cultural meaning at the aggregate level. The procedures can reliably identify cultural similarities and differences among social groups in the degree of sharing and in the meaning of specific elements of cultural domains. Hypotheses about between-group differences in culture or in the association of cultural factors and other variables can then be evaluated (for examples, see Kempton, Boster, and Hartley 1996; Caulkins 2001; Jaskyte and Dressler 2004).

The cultural consensus model provides one link from the aggregate level of shared meaning to the individual level—the cultural competence coefficient—and for testing certain kinds of hypotheses, this measure will suffice; however, people do not just know or think things, they do and believe things, and
assessing the degree to which individuals conform in their behaviors and their personal beliefs to cultural prototypes for those behaviors and beliefs is an important question (Crossley 2001). We have found the expression of the cultural in the individual to be important in research in medical anthropology. In previous research in Brazil, we suggested that cultural consonance, or the degree to which individuals in their own behaviors approximate the prototypical behaviors encoded in a cultural model, might be related to health outcomes. We found that higher cultural consonance in two different domains (lifestyle and social support) was associated with lower arterial blood pressure, lower perceived stress, fewer symptoms of depression, and a higher sense of one’s own efficacy (Dresser and dos Santos 2000; Dressler, Balieiro, and dos Santos 2002). This research suggests an approach to measurement in anthropology.

MEASURING CULTURAL CONSONANCE

Our previous approach to measuring cultural consonance in a single domain (lifestyles) is described succinctly in Dressler (1996). This was a rudimentary approach, building on years of previous research in which a scale to measure lifestyle (material goods and related behaviors) had been developed using procedures similar to Janes’s (1990). Applying the cultural consensus model to those data, we tested to see if our ethnographic insights had been correct.

Our current research seeks to refine and systematize this approach in three ways. First, drawing on the strongest qualitative and quantitative methods in cognitive anthropology, we improved our description of the cultural models for behavior. Second, we made better use of this description in developing measures of cultural consonance. Third, we expanded the cultural domains examined to understand better the measurement of cultural consonance in domains with different characteristics. The domains examined include lifestyles, social support, family life, and national characteristics. The measurement of cultural consonance will be described for each of these domains.

We used the tools of cognitive anthropology for collecting data (Weller and Romney 1988). The specific approach in each domain are described below. In general, we used an iterative approach of collecting data about the domain with free lists, pile sorts, and rating and ranking tasks. Specific steps depended on the domain under study. We then proceeded to test for cultural consensus. This last step was also designed to create the measure of cultural consonance in the survey that followed.
We did not rely exclusively on these semistructured interview techniques. To triangulate our understanding of these domains, we also used focused group interviews and individual unstructured interviews. The aim of data collection here was to determine if similar terms and semantic relations emerged from these minimally directed interviews. Discussion of these data is beyond the scope of this article; however, in general, the results are consistent with the results from the more structured cultural domain analysis.

With respect to sampling, Handwerker and his associates (Handwerker, Harris, and Hutcherson 1997; Handwerker and Wozniak 1997) have argued that in cultural domain analyses, strict canons of statistical sampling to achieve independence of cases do not apply, given the underlying assumption that cases are not independent in any event since they share cultural models. This view was adopted here, with the caveat that, especially in a society as diverse and complex as Brazil, sampling potential intracultural diversity is important. Therefore, when respondents were recruited for various phases of cultural domain analyses, stratified sampling was used. Equal groups of men and women were recruited, as were equal groups of persons younger and older than 45 years. We included equal groups of persons with primary, secondary, and university levels of education. The sample sizes for different steps are as follows: free lists \((n = 43)\), pile sorts \((n = 16)\), pile sorts/rankings \((n = 34)\), rankings \((n = 22)\), and cultural consensus analysis \((n = 66)\).

Conventional psychometric characteristics of the measures were assessed using data from an epidemiologic survey following the cultural domain analysis. Four neighborhoods varying in socioeconomic status were selected. These are the same four neighborhoods used in our research ten years ago, described in other articles (Dressler, Balieiro, and dos Santos 1997). In the current study, households were randomly selected from complete listings of occupied addresses within each neighborhood. Both heads of household (if present) and one child older than 18 years were invited to participate in the research. In 60% of households contacted, at least one individual agreed to participate, and 71.2% of households contributed more than one respondent. The final sample size was 271 individuals.

**Lifestyle**

Lifestyle refers to the material accouterments and behavioral manifestations of being a success in life (Bourdieu 1984). Understanding the cultural model of lifestyle began with breaking the domain into two subdomains: material goods and leisure activities. This was done to facilitate the interview process. In the free list, respondents were asked to list material goods or pos-
sessions that people need to live a good life. For leisure activities, they were asked to list the activities in which people typically engage in their free time.

The free list sample generated eighty material goods and sixty-six leisure activities. From those lists, twenty-one items from each subdomain were selected for further study. Research staff together selected items occurring higher in the lists of more respondents (a criterion referred to as “salience”). Some items lower in the lists were included to ensure that the full range of semantic variation within the domains was sampled.

Within each domain, another sample performed an unconstrained pile sort of the terms. Each of the terms was written on an index card, and respondents grouped the terms on the basis of similarity. They were instructed to make as many piles as they wished. Throughout the task, the comments of the respondents were noted, and at the end of the task, each respondent was asked to explain why he or she had made those particular groupings. Using Anthropac 4.05 (Borgatti 1993), the pile sorts were converted to aggregate proximity matrices and analyzed with nonmetric multidimensional scaling (MDS; Kruskal and Wish 1978, again in Anthropac) so that similarities and differences in meaning among the terms could be visualized in two dimensions. Adequate fit was obtained in two dimensions in each subdomain (stress = .15). (Space limitations preclude reproducing the visual representations.)

For material possessions, respondents’ pile sorts were dominated by ideas of what you really need for a good life versus what is superficial. They attended to a single evaluative dimension, with “need” being the operative attribute. For leisure activities, respondents attended to two attributes of the activities. One was a sense of personal development resulting from the activity (e.g., reading or studying). The other attribute was social interaction (e.g., going to bars, conversing with friends). These hypothesized dimensions were explored in the next rounds of interviews and were found to account for the similarities and differences in meaning of the elements in each subdomain using property-fitting analysis (PROFIT, described by Kruskal and Wish [1978] and available in Anthropac).

At this point, we examined cultural consensus on the principal dimension of lifestyle that would be essential for calculating cultural consonance in lifestyle. This was the importance of the items “to live,” a phrase that sounds ironic in English but works well in Portuguese to describe what are regarded as those elements of a lifestyle that no one should have to live without to have a decent life. In the final consensus interview, respondents were asked to rate, on a four-point scale ranging from not at all important to very important, the importance of an item for having a life. At no point were respondents asked
about their own lifestyles. They were instructed to evaluate these items in terms of what was generally thought to be important in the community.

A combined list of thirty-three material goods and leisure activities was rated. Using the cultural consensus model as derived for rank data (Romney, Batchelder, and Weller 1987), there was high consensus among the respondents (ratio of the first-to-second eigenvalue = 6.59, mean competence = .71 ± .12). When intracultural variation was examined in the distribution of competence, significant differences were found between the education groups $(p < .01)$, with the least well-educated respondents having the highest competence. Interestingly, the more well-educated respondents were less convinced of the importance of certain items than were the less well-educated respondents, but the differences were not substantial enough to suggest more than one cultural model. It should also be noted that when combined as a single group of items, respondents had no trouble thinking of material goods and leisure activities along a single evaluative dimension. These items described a coherent domain of lifestyle.

To assess cultural consonance in lifestyle, in the survey, individuals were asked to indicate whether they owned each material good. On the leisure activities, individuals reported on a four-point scale (ranging from never to several times per week) the frequency with which they engaged in those behaviors. Two ways of calculating cultural consonance in lifestyle can be used with these data. The first is to use all thirty-three items, weighting each item by the consensus ratings from the answer key in cultural consensus analysis. Doing this yields a normally distributed scale with an internal consistency reliability of alpha = .82.

The other way to calculate cultural consonance in lifestyle is to limit the selection of behavioral items to those items rated in the cultural model as being at least “important” in having a life. This reduces the number of items to nineteen. Then, the reporting of these items can be counted, and the proportion of items of importance reported by the respondent can be calculated. This scale is also approximately normally distributed, although the internal consistency reliability is lower (alpha = .67, in part a function of the fact that many persons possess these more important items). This is an adequate reliability, and, it turns out, this reduced scale has slightly higher correlations with outcome variables (see below). Therefore, this latter measure was retained, and it is shown in Table 1.

Social Support

For social support, we collected two free lists. The first was a list of problems for which people typically seek the help of others (fifty-five terms). The
second was a list of the kinds of people to whom one might turn for help (thirty-five terms). We reduced this to eight problems and seven potential supporters.

As the free lists were collected, it was apparent that people thought in terms of social support as a hierarchy of resort within each problem type (e.g., when confronted with debt, people would first ask one type of person, proceeding next to another, and so on). This appeared so uniform in the interviews that no further exploration of the domain was necessary. We returned to the domain of social support only in the final cultural consensus analysis.

In the interview for cultural consensus analysis, each respondent was presented with seven cards on which the names of potential supporters were written. They were then presented with a problem and asked to rank the order in which they thought it was typical for people to ask different kinds of people for help. When analyzed for consensus, there was substantial agreement

### TABLE 1

Cultural Consonance in Lifestyle

<table>
<thead>
<tr>
<th>Item</th>
<th>Proportion^a</th>
<th>Rating of Item Importance in Consensus Model^b</th>
<th>Rank of Importance in Consensus Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. House</td>
<td>.80</td>
<td>3.92</td>
<td>1.5</td>
</tr>
<tr>
<td>2. Stove</td>
<td>1.00</td>
<td>3.92</td>
<td>1.5</td>
</tr>
<tr>
<td>3. Personal study</td>
<td>.43</td>
<td>3.86</td>
<td>3.0</td>
</tr>
<tr>
<td>4. Money for school</td>
<td>.55</td>
<td>3.84</td>
<td>4.0</td>
</tr>
<tr>
<td>5. Refrigerator</td>
<td>.99</td>
<td>3.81</td>
<td>5.0</td>
</tr>
<tr>
<td>6. Time to rest</td>
<td>.61</td>
<td>3.79</td>
<td>6.0</td>
</tr>
<tr>
<td>7. Talk with friends</td>
<td>.79</td>
<td>3.70</td>
<td>7.0</td>
</tr>
<tr>
<td>8. Telephone</td>
<td>.88</td>
<td>3.69</td>
<td>8.0</td>
</tr>
<tr>
<td>9. Time to read</td>
<td>.65</td>
<td>3.62</td>
<td>9.0</td>
</tr>
<tr>
<td>10. Play sports</td>
<td>.37</td>
<td>3.61</td>
<td>10.0</td>
</tr>
<tr>
<td>11. Sofa</td>
<td>.90</td>
<td>3.44</td>
<td>11.5</td>
</tr>
<tr>
<td>12. Dining table</td>
<td>.96</td>
<td>3.44</td>
<td>11.5</td>
</tr>
<tr>
<td>13. Go to church</td>
<td>.48</td>
<td>3.36</td>
<td>13.0</td>
</tr>
<tr>
<td>14. Money for extras</td>
<td>.53</td>
<td>3.32</td>
<td>14.0</td>
</tr>
<tr>
<td>15. Car</td>
<td>.70</td>
<td>3.16</td>
<td>15.0</td>
</tr>
<tr>
<td>16. Television</td>
<td>.96</td>
<td>2.98</td>
<td>16.0</td>
</tr>
<tr>
<td>17. Computer</td>
<td>.42</td>
<td>2.98</td>
<td>17.0</td>
</tr>
<tr>
<td>18. Washing machine</td>
<td>.67</td>
<td>2.90</td>
<td>18.0</td>
</tr>
<tr>
<td>19. Web access</td>
<td>.40</td>
<td>2.75</td>
<td>19.0</td>
</tr>
</tbody>
</table>

^a For items 1, 2, 4, 5, 8, 11, 12, 14, 15, 16, 17, 18, and 19, this refers to the percentage possessing the item. For items 3, 6, 7, 9, 10, and 13, this refers to the proportion reporting that they engage in these activities at least once per week.

^b Based on the rating of not at all important = 1 to very important = 4.
These rankings (ratio of the first-to-second eigenvalue = 6.53, mean competence = .67 ± .14). The consensus rankings are shown in Table 2.

To measure cultural consonance in social support, respondents in the survey sample were provided with the same set of cards for potential supporters and the same problems but were asked to rank only the first three or four of their personal choices for support. This was done purely for practical reasons because in pretesting the survey interview, we found that this task slowed the flow of the interview substantially. To calculate cultural consonance in social support, we first transposed the data matrix so that each respondent became a column and each row was that respondent’s ranking of a particular supporter for a particular problem. Supporters not ranked by the respondent were assigned the mean of the missing ranks. A column was added that was the consensus ranking of a particular supporter in relation to a particular problem. We could then calculate a simple correlation coefficient between the rankings by each respondent and the rankings from the consensus analysis. This correlation is used as the measure of cultural consonance in social support. This measure ranges from −.25 to .81, with a mean of .49 (±.19). The distribution is slightly skewed to the right, but not substantially so.

### Family Life

In the free list, we posed two questions to respondents. First, we asked them to imagine a family they admired and to tell us the characteristics of that
family. Second, we asked them to imagine a family they did not admire and to
tell us the characteristics of that family. This resulted in lists of eighty-nine
and ninety-two terms, respectively. We consolidated and reduced this list to
twenty-four terms.

We then used a constrained pile sort. Respondents were asked to create
two piles, those terms characteristic of good families and those terms charac-
teristic of bad families, since this difference was already incorporated in the
lists of items. Within each pile, they were free to create as many piles as they
wished, the only requirement being that they had to create at least two piles.
The MDS (stress = .04) and cluster analysis of these data suggested a single
evaluative dimension separating characteristics of good versus bad families
(and this was confirmed in a PROFIT analysis, using rankings collected
later). There were category differences at either end of the continuum. The
category difference for positively evaluated terms separated terms referring
to family structure (e.g., organized) from terms referring to the affective cli-
mate of the family (e.g., love, understanding). The category difference for
the negatively evaluated terms separated characteristics such as bad manners
(e.g., disrespect, egoism) from characteristics such as violence and substance
abuse (e.g., violence, addiction).

In the consensus interview, we reduced the number of items to thirteen,
primarily because of concerns about the influence of the ends of the
 evaluative continuum; that is, were concerned that consensus could be
generated by agreement about what goes at the ends of the continuum (e.g.,
people agreeing that “love” is good but “violence” is bad) and not about how
elements are arrayed in between. We eliminated the most negatively evalu-
ated items and sampled from the most positively evaluated items. In the con-
sensus sample, respondents were asked to rank these in terms of their impor-
tance “in order to have a family.” There was a high degree of consensus on
this ranking (ratio of the first-to-second eigenvalue = 7.42, mean competence =
.82 ± .09). Despite the very high overall consensus on the ranking of these
family characteristics, we can still detect a small but statistically reliable dif-
fERENCE in competence between men and women (.79 vs. .85, p < .01). This
indicates that women exchange and share the meaning of family concepts
more than men do.

The challenge at this point was locating individuals in the space of mean-
ing defined by these terms. In other cultural domains, it was a straightforward
matter to translate the culturally salient items into questions about individual
behavior. The domain of family life presented a different challenge. Is it pos-
sible to ask people if, for example, their family is well organized? Or if their
family members really love each other? We believed responses to such ques-
tions would be dominated by social desirability. So, we assessed perceptions of family life.

We presented individuals with a statement about the family, phrased explicitly in terms of their own family, and asked them to agree or to disagree with that statement. For each concept, we generated a statement describing the family that would be acceptable in everyday speech. We generated a minimum of one sentence for each concept, more important concepts having two statements. For some of the concepts, acceptable phrasing in Portuguese enabled us to use simple statements (e.g., “In my family we feel close to one another”). For some concepts, linguistically it was better to phrase the statement in terms of a wish or desire (e.g., “At times I wish my family was more organized”). All items were written in Portuguese.

To calculate a total scale score, we weighted individual responses by the importance of that particular concept in the cultural consensus analysis. For example, if a person strongly agrees that there is a great deal of love in their family, they receive more points for this than if they strongly agree that their family firmly confronts life’s problems, because in the consensus analysis, “love” was seen as more important than the concept of firmeza (firmness). The weights were adapted from the consensus rankings of the importance of the items. We reduced the thirteen ranks to five weights because similar items had similar consensus ranks. The scale has quite high internal consistency reliability (alpha = .89) and is shown in Table 3.

National Characteristics

The free list for national characteristics started with the question, “What characteristics are most important in defining a Brazilian?” This query generated 133 distinct terms, of which 26 were retained for further analysis. Analysis of the pile sorts of these characteristics had a good fit in two dimensions (stress = .05) and clearly distinguished items regarded as negative characteristics of Brazilians from those considered positive. Discussions by respondents during the pile sorts indicated that this was a contested cultural domain since some Brazilians marveled at how accurately we captured the Brazilian character, while others were adamant that these were scurrilous stereotypes. Positive characteristics included terms such as hard workers, happy, and receive others well, while negative terms included take advantage, lazy, ignore the poor, and corrupt.

In the consensus interview, respondents rated the terms on a four-point scale ranging from total disagreement that this represented a characteristic of Brazilians to total agreement that this represented a characteristic of Brazilians. There was a modest consensus on this rating (ratio of the first-to-second
eigenvalue = 3.97, mean competence = .57 ± .19). The items rated as most characteristic of Brazilians included a mix of the positive and negative items.

Like the domain of family life, we wrote items representing personal beliefs in this domain and asked respondents to agree or disagree with the statements (again, all items were generated in Portuguese). With the survey data, we examined a unidimensional scale of all items, but this had very low internal consistency. Through a series of exploratory factor analyses, items with low communalities were deleted. It became apparent that the items referring to the most positive characteristics of Brazilian life had such low variability that they did not covary with other items (i.e., people believe consistently in positive descriptors). There was, however, a consistent factor that combined most of the more unfavorable aspects of Brazilian

### TABLE 3

<table>
<thead>
<tr>
<th>Item Model</th>
<th>Weight from Consensus Model</th>
<th>Item Mean from Survey Data</th>
<th>Item-Total Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In my family, we feel close to one another.</td>
<td>4</td>
<td>2.50</td>
<td>.45</td>
</tr>
<tr>
<td>Sometimes I wish my family were more organized.</td>
<td>3</td>
<td>2.15</td>
<td>.32</td>
</tr>
<tr>
<td>At times when I need it, there is no one to help resolve problems.</td>
<td>2</td>
<td>1.04</td>
<td>.35</td>
</tr>
<tr>
<td>People in my family are hard workers.</td>
<td>3</td>
<td>2.70</td>
<td>.32</td>
</tr>
<tr>
<td>At times, we avoid one another.</td>
<td>4</td>
<td>0.86</td>
<td>.69</td>
</tr>
<tr>
<td>At times in my family, I wish we felt more love for one another.</td>
<td>5</td>
<td>1.62</td>
<td>.49</td>
</tr>
<tr>
<td>We are as well adjusted as a family could be.</td>
<td>3</td>
<td>2.05</td>
<td>.66</td>
</tr>
<tr>
<td>When I do something, I don’t think about my family.</td>
<td>1</td>
<td>0.76</td>
<td>.22</td>
</tr>
<tr>
<td>I think my family criticizes too much.</td>
<td>1</td>
<td>1.08</td>
<td>.54</td>
</tr>
<tr>
<td>My family firmly confronts problems.</td>
<td>2</td>
<td>2.26</td>
<td>.46</td>
</tr>
<tr>
<td>Normally, mine is a happy family.</td>
<td>3</td>
<td>2.31</td>
<td>.61</td>
</tr>
<tr>
<td>We understand each other completely.</td>
<td>4</td>
<td>2.17</td>
<td>.71</td>
</tr>
<tr>
<td>We help each other with problems.</td>
<td>2</td>
<td>2.33</td>
<td>.61</td>
</tr>
<tr>
<td>We don’t have time to listen to each other.</td>
<td>4</td>
<td>1.05</td>
<td>.46</td>
</tr>
<tr>
<td>At times, we don’t have sufficient respect.</td>
<td>1</td>
<td>0.82</td>
<td>.66</td>
</tr>
<tr>
<td>I can talk about important things in my family.</td>
<td>4</td>
<td>2.32</td>
<td>.62</td>
</tr>
<tr>
<td>We feel love for one another.</td>
<td>5</td>
<td>2.38</td>
<td>.67</td>
</tr>
<tr>
<td>At times, I wish my family didn’t fight so much.</td>
<td>1</td>
<td>1.45</td>
<td>.41</td>
</tr>
</tbody>
</table>

a. In the consensus model, items were ranked 1 to 13, but the consensus ranks could be reduced to 5.
b. In the survey, participants’ responses ranged from disagree totally (0) to agree totally (3).
c. These items were reversed in direction prior to scoring.
life. These items are shown in Table 4. These eight items have acceptable internal consistency reliability (alpha = .69) and represent what we have come to think of as “cultural cynicism.” That is, those individuals who endorse more of the items have a more cynical view of Brazilians and Brazilian life, but it is a distinctly culturally constructed cynicism.

Construct Validity

Messick’s (1995) perspective on construct validity is used here: “Validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment” (p. 741).

Messick (1995) suggests six aspects of construct validity. These are (1) the content aspect (or face validity), (2) the substantive aspect (the theoretical rationale for the measure), (3) the structural aspect (how well the scoring system reflects the distribution of the attribute), (4) the generalizability aspect (how well the measure functions across populations), (5) the external aspect (convergent-discriminant validity and criterion relevance), and (6) the con-
sequential aspect (the value implications of score interpretations for social action).

Messick’s (1995) first, second, and fifth aspects of construct validity are relevant to evaluate cultural consonance measures. With respect to content, the measures of cultural consonance were generated directly from informants’ words, after testing for cultural consensus. It is difficult to imagine a procedure that would retain more fidelity with respect to local meaning and understanding. With respect to the substantive aspect, the measures are embedded in a theory of cultural models. Furthermore, various theorists have suggested the value of measuring what we term cultural consonance (e.g., Sapir 1946).

Messick’s (1995) fifth aspect of construct validity can be examined in three ways. The first is to examine the differences across the four neighborhoods that vary socioeconomically. Brazil is one of the most highly stratified societies in the world (Rezende 1998), and access to economic resources should affect the ability of individuals to act on at least some of these cultural models, hence limiting those individuals’ cultural consonance. These results are shown in Figures 1 and 2. In Figure 1, both cultural consonance in social support and cultural consonance in lifestyle differ significantly across the four neighborhoods, with the wealthiest neighborhood showing the highest consonance \( p < .001 \). In Figure 2, cultural cynicism differs across the four neighborhoods \( p < .001 \), but cultural consonance in family life does not. Persons in the wealthiest neighborhood harbor the least culturally cynical views, and there is no socioeconomic level in which persons perceive their families as more (or less) consonant with the prototypical model.

A second approach is to examine the correlations of cultural consonance with several psychological variables generally thought to be a part of the process of psychosocial stress. The cultural consonance model was developed in the context of research on health, and in general, we can regard low cultural consonance as a chronically stressful experience. The correlations of the cultural consonance variables with measures of psychological stress are shown in Table 5. With the exception of cultural consonance in social support, all of the consonance measures are associated with the psychological variables in a predictable direction. That is, higher cultural consonance is associated with less stress and depression and greater internal locus of control. The exception to this is cultural cynicism, which is associated with an external locus of control and more stress and depression.

Finally, the pattern of correlations among the cultural consonance variables can be thought of as a measure of construct validity. In general, being higher on one measure is indicative of being higher on another measure,
again with the understandable exception of cultural cynicism. Higher cultural cynicism is associated with a lower cultural consonance on all other variables. These four cultural consonance variables load together on a single principal component that accounts for 46% of the variance. The loadings on the principal component are as follows: cultural consonance in lifestyle (.781), cultural consonance in social support (.617), cultural consonance in family life (.513), and cultural cynicism (−.765). There is a single continuum that represents a “general cultural consonance.” At one end are persons with higher cultural consonance in lifestyle, social support, and family life and a lower cultural cynicism; at the other end are people with a higher cultural cynicism and lower cultural consonance in lifestyle, family life, and social support. This general cultural consonance factor also is significantly associated with the psychological outcome variables (see Table 5).
DISCUSSION

The aim of this article was to present and evaluate a measurement model for anthropology. This model draws on and incorporates criteria for evaluating measures as those criteria have been developed in related fields, principally psychology. Primarily, however, this model depends on a more systematic and transparent set of procedures for getting from an assessment of the collective knowledge that individuals share to the measurement of individual behavior and personal belief.

Adequate measurement in anthropology depends on the degree to which it reflects collective meaning. Traditionally, demonstrating appropriateness of a measure has depended on the rhetorical skill of the researcher. The mea-
The measurement model presented here clarifies that process. Collective meaning can be evaluated using the methods of cognitive anthropology and especially the technique of cultural consensus analysis. A theory of the health effects of cultural consonance then requires that the instantiation of that collective meaning in individual behavior be assessed. The set of procedures outlined here draws a clear line from collective meaning to individual behavior. A claim that the measures of cultural consonance presented here are measuring that which they are intended to measure is less ambiguous.

The measurement model presented here is also flexible. In the cultural domains examined, three different assessments of cultural consonance have been obtained. The first, and perhaps the most straightforward, is exemplified by cultural consonance in lifestyle. It is a simple step to assess the extent to which an individual’s (reported) behavior matches the collective evaluation of lifestyle.

The measurement of cultural consonance in social support uses a similar approach. Individuals report their behavior, but it is clearly hypothetical. Consonance is measured as a correlation of their individual profile with the collective profile. In many respects, this is close to a cultural competence coefficient, except that individuals are reporting their own behaviors, not their knowledge of patterned behavior. This might be thought of as a hybrid measure of competence and consonance.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cultural consonance in lifestyle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Cultural consonance in social support</td>
<td>.343**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Cultural consonance in family life</td>
<td>.141*</td>
<td>.186*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Cultural cynicism</td>
<td>−.485**</td>
<td>.202**</td>
<td>.269**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. General cultural consonance factor</td>
<td>.781**</td>
<td>.617**</td>
<td>.513**</td>
<td>.765**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Locus of control</td>
<td>.397**</td>
<td>.104</td>
<td>.244**</td>
<td>.413**</td>
<td>.444**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Perceived stress</td>
<td>−.293**</td>
<td>.015</td>
<td>−.261**</td>
<td>.242**</td>
<td>.255**</td>
<td>−.356**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Depressive symptoms</td>
<td>−.316**</td>
<td>.150*</td>
<td>−.214**</td>
<td>.316**</td>
<td>.375**</td>
<td>−.382**</td>
<td>.641**</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
The measurements of cultural consonance in family life and cultural cynicism diverge most from previous measures of cultural consonance because here we are evaluating the degree to which individuals in their personal beliefs about the world correspond to collective meaning. Some researchers use the terms knowledge and belief as virtually interchangeable, but for our purposes, it is better to distinguish these concepts. The scheme presented can be viewed in Searle’s (1964) sense of a constitutive rule. This is a rule that defines an “x” as an “x” and not as a “y,” a “z,” or a “q.” Therefore, to “know” what marriage is in the American kinship system is to know something to the effect that one man and one woman form a lifelong commitment sanctioned by codified laws involving exclusive sexual access, pooling of economic resources, and socialization of common children. To “believe” something about American marriage involves an evaluation of that definition or parts of it. So, the measurement of cultural consonance in family life and cultural cynicism assess what individuals personally believe about those domains, not what they know.

Cultural domain analysis and cultural consensus analysis enable the researcher to discover and describe the various cultural spaces inhabited by their respondents. The tools of psychometric theory then enable the researcher to evaluate how well he or she can locate individuals in those cultural spaces. The various measures of cultural consonance appear satisfactory in this regard. All of the measures for which conventional criteria of reliability can be calculated have adequate reliability. Three approaches to the assessment of construct validity suggest that these measures of cultural consonance do indeed measure what they purport to measure.

A variety of directions could be taken in future work on this model. First, it would be interesting to try to build into the measurement of cultural consonance the fact that cultural models in some domains are strongly shared (e.g., lifestyle), while in other domains, the cultural model seems more highly contested (e.g., national characteristics). Second, a variety of approaches could be used to construct weights for items in the scales of cultural consonance. For example, “person fit theory” in psychometrics is an effort to identify the reasons why some individuals do not match an anticipated pattern on a particular test. In lieu of a theoretically defined pattern, person-fit analyses make use of the observed distribution in a sample, weighting the importance of items by the proportion of individuals answering them correctly (Meijer and Sijtsma 1995). Coupled with the weights from cultural consensus analysis, adaptation of person-fit procedures would add an additional dimension to the construction of a scale. Fourth, it would be interesting to examine cultural consonance along multiple dimensions of the cultural models. Because of
our specific interests, the “importance” (a general evaluative dimension) of items was the most appropriate dimension for constructing a measure of cultural consonance. But certainly, most cultural models are constructed out of multiple dimensions of meaning, and the way in which individual profiles of belief and behavior match these multiple dimensions of meaning would be interesting. Finally, in future work with this measurement model, it would be useful to evaluate its applicability in the evaluation of various hypotheses. In this regard, the concept of cultural consonance may be sufficiently general to be applied outside of medical anthropology, in which it was developed. It may be that the instantiation of the cultural in individual behavior and personal belief can be a useful measure for other questions (see, for example, Chick’s [1981] work on the civil-religious hierarchy in Mesoamerica). The only way to further evaluate and to extend this measurement model, however, is replication.

NOTES

1. Space precludes a careful description of the community, which would help the reader to appreciate better the measurement process (note that we are not leaving traditional ethnography behind). Detailed descriptions of the community can be found in Dressler, Balieiro, and dos Santos (1997) and Dressler et al. (2004).

2. Handwerker’s argument is not about the need for independence in measurement; rather, he is referring to lack of independence in cases due to the sharing of cultural models. Data must still be collected independently.

3. We became embroiled in a long discussion about using written materials because of the potential low literacy levels at the lower end of the socioeconomic continuum. We decided that our least well-educated respondents were sufficiently literate to recognize words, at least with assistance from the interviewer. Our confidence in this procedure was bolstered by the fact that one of the research staff members also taught adult literacy classes.

4. Formal process models for cultural consensus analysis have been derived for dichotomous and multiple-choice formats of data collection; informal data models have been developed for rank-order and interval-level formats (Romney and Batchelder 1999). Obviously, the data used here in two domains (lifestyle and national characteristics) fit neither of these data formats, in that they are collected as ordered polychotomies, while two data formats (social support and family life) fit the rank-order data formats. We used the informal data model of cultural consensus analysis developed for rank-order and interval-level data for the ordered polychotomies, with three rationales. First, some researchers argue explicitly that cultural consensus analysis can be generalized to these types of data (Kempton, Boster, and Hartley 1995; Ross 2004:137–48). Second, we asked our questions in a specific way. When respondents were presented with items they were asked if they agreed or disagreed that the item was, for example, important for having a good life. They were then asked to qualify the strength of their agreement (or disagreement). This meant that we could dichotomize the responses as agree-disagree and use the formal process model of cultural consensus theory for dichotomous data; there were no important differences in
the results between this analysis and the informal data model for rank data. And third, in a separate subsample, we collected rank-order data in the domain of lifestyle; there were no differences in the results of this analysis compared to the data collected as ordered polychotomies.

Obviously, we are not trying to break new ground in the mathematical derivation of cultural consensus analysis here; rather, we are trying to use the technique to further our investigation of a substantive issue (in this case, individual psychophysiologic adaptation). In doing so, we are applying the model to data that, while not known to be completely appropriate for cultural consensus analysis, are particularly well suited for collecting informants’ ideas about certain domains. Our informants were most comfortable with the use of ordered polychotomies in the domains of lifestyle and national characteristics because it allowed them to express a qualified certainty about the importance of items. This may violate the strict mathematical requirements of the model, but in the long run, given the convergent evidence regarding the accuracy of the results, we believe that this violation does little damage to the validity of our inferences.

5. Globally perceived stress (S. Cohen, Karmack, and Mermelstein 1983) and depressive symptoms (Silveira and Jorge 2000) are measures of the subjective experience of stressful circumstances. Locus of control (Coreil and Marshall 1982) is a measure of the sense that one is in control in one’s life. All these measures have adequate reliability in this sample (alphas > .70). The depressive symptom scale is a Portuguese translation of the Center for Epidemiologic Studies—Depression Scale, adapted and validated by Brazilian researchers. Cohen’s perceived stress scale is a widely used measure of perceived stress. The locus of control scale was developed by Coreil and Marshall (1982). We used the latter two scales in an earlier study in Brazil in which they both were translated and back-translated to ensure comparability (Dressler, Balieiro, and dos Santos 2002).

REFERENCES


**WILLIAM W. DRESSLER** is a professor of anthropology in the Department of Anthropology, University of Alabama, and a College of Arts & Sciences Leadership Board Faculty Fellow. His research interests are in the area of medical anthropology and, more specifically, cultural influences on disease risk. Recent publications include “What’s Cultural about Bio-cultural Research” (*Ethos* 2005) and “Cultural Consonance and Arterial Blood Pressure in Urban Brazil” (*Social Science and Medicine* 2005, with R. Ribeiro, M. Balieiro, and J. E. dos Santos).

**CAMILA D. BORGES** received her M.A. in psychology from the University of São Paulo-Ribeirão Preto. Her research and clinical interests are in the area of the family. Recent publications include “Deficiência Auditiva: Escolarização e Aprendizagem e Língua de Sinais Na Opinião Das Mães” (*Paidéia: Cadernos de Psicologia e Educação* 2002, with E. Petean) and “Maturidade Emocional, Locus de Controle e Ansiedade em Adolescentes Obesos” (*Paidéia: Cadernos de Psicologia e Educação* 2001, with A. Cavalho, J. Netto, and others). She practices clinical psychology in São Simão, Brazil.

**MAURO C. BALIEIRO** is a Ph.D. candidate in psychology at the University of São Paulo-Ribeirão Preto, a professor at Paulista University, and a psychologist in private practice in Ribeirão Preto. His research interests are in the area of cultural influences on mental health. Recent publications include “Cultural Consonance and Arterial Blood Pressure in Urban Brazil” (*Social Science and Medicine* 2005, with W. Dressler, R. Ribeiro, and J. E. dos Santos) and “Cultural Consonance and Psychological Distress” (*Paidéia: Cadernos de Psicologia e Educação* 2002, with W. Dressler and J. E. dos Santos).
JOSÉ ERNESTO dos SANTOS is an associate professor of medicine on the Faculty of Medicine of Ribeirão Preto, University of São Paulo-Ribeirão Preto. He is a specialist in lipid metabolism and the treatment of hyperlipidemia and weight disorders. His research interests include cultural and psychological influences on health. He and William W. Dressler have collaborated for more than twenty years on research on social and cultural influences on health in urban Brazil, summarized in many joint publications including “Social and Cultural Dimensions of Hypertension in Brazil: A Review” (Cadernos de Saúde Pública 2000) and “Correlações Sociais e Culturais com a Pressão Arterial: Os Estudos de Dressler e Dos-Santos” (Revista Brasileira de Hipertensão 2001).