

CULTURAL CONSENSUS THEORY

Susan C. Weller

University of Texas Medical Branch

Galveston, Texas

2011

WORKSHOP OVERVIEW

I. Overview of Cultural Consensus Theory (Purpose)

II. Developing Interview materials

III. Selecting informants

IV. Estimating Cultural Beliefs

V. Examples & Interpreting Results

PURPOSE

- Cultural consensus theory can be used to test whether beliefs are shared among community members
- What are the culturally appropriate answers to a series of questions?
- How much does each person know?
- It is different from standard knowledge tests, where the answers are known beforehand, and informants' responses are scored as correct/incorrect.
- Also, consensus is not created; agreement is measured, to see if consensus is present.

INTERVIEW MATERIALS: A Set of Related Questions

The first step in a study of cultural beliefs, is the development of interview materials appropriate for those beliefs.

Questions may be developed from information obtained from a variety of sources:

New items may come from open-ended, structured interviews with individuals or small groups. (Sample should be similar to those you wish to study)

Items may be taken from existing sources, such as scientific publications, archival records, questionnaires and/or tests.

INTERVIEW MATERIALS: A Set of Related Questions

- I. To develop interview materials, themes/items culled from:
 1. Open-ended, in-depth interviews with individuals
 2. Open-ended, in-depth interviews with groups.
 3. Existing sources, such as scientific publications & archival records
 4. From standardized questionnaires, tests and/or scales.

- II. To design an structured interview for detailed study of variation:
The goal is to obtain 20 or more questions all on a single topic
all in the same format.

A Set of Related Questions

If we wanted to learn about people's beliefs about AIDS, we would begin by interviewing a sample of people about their explanatory model of AIDS.

Phase I: A first set of interviews to collect relevant themes
Open-ended questions would be used to elicit perceived causes, symptoms, and treatments for AIDS.

Themes from those interviews and from other sources would then be used to create a set of questions designed to systematically study beliefs about AIDS.

Phase II: A second set of interviews to ask each person about all of the themes.

A Set of Related Questions

Phase I: Descriptive interviews collect relevant themes in explanatory models of AIDS.

Phase II: Responses from those interviews (and other sources) are used to create a set of systematic questions:

For example:

Is AIDS inherited?

Can you get AIDS from a public bathroom?

Can you be infected with AIDS when you donate blood to someone else?

Are people with AIDS more susceptible to getting other illnesses?

Is a cold that won't go away a symptom of AIDS?

Do you have diarrhea with AIDS?

If you have a positive attitude, can you help cure AIDS?

Is a doctor the best person to treat AIDS?

What Kind of Questions in Phase 2?

1. Open-ended questions, only if the answer is a single word or short phrase.
2. Multiple choice type responses:
 - a. Dichotomous: true/false or yes/no questions.
 - b. Multiple choice: morning/afternoon/evening
3. Ranked or interval type responses:
 - a. Order these cards from most to least important
 - b. How far is it from A to B?
 - c. Rating scales for each item

A Set of Related Questions

1. Open-ended single word answer questions
2. Multiple choice type questions:
Must be balanced
3. Ranked or interval type questions:
Rating scales must be balanced

THE GOAL IS TO HAVE 20+ QUESTIONS

All on the same topic

ALL IN THE SAME FORMAT

SUMMARY (Questions)

1. Open-ended, exploratory interviews collect important themes
2. Results used to create structured interview

20+ questions: All on the same topic,
All in the same format

Estimating Cultural Beliefs

How can we identify and measure cultural beliefs?

Some have debated how to describe culture in the face of some degree of heterogeneity.

D' Andrade (1987) sums it up nicely:

“For a long time there has been a minor scandal at the heart of the study of culture. This scandal goes as follows: All human groups have a culture; culture is shared knowledge and beliefs; but when we study human groups, we find that there is considerable disagreement concerning most items of knowledge and belief.

Cultural Beliefs

...One resolution of this problem is to treat the most frequently held items of knowledge and belief – the modal items – as the culture of the group. No doubt more than one living middle class Southern Californian believes in witchcraft. But the fact that someone holds this belief does not make it part of current middle-class Californian culture – if one says that culture consists of modal items.”

D' Andrade (1987)

STATISTICAL TERMINOLOGY

Measurement level

1. Categorical or nominal
2. Ordinal or ranked
3. Interval measurement
4. Ratio

Univariate statistics

1. Mode = most frequently occurring category
2. Mean = mathematical average or weighted average
3. Standard deviation = dispersion around mean

Cultural Beliefs

How can we identify and measure cultural beliefs?

Cultural beliefs are learned and shared beliefs.

Aggregate responses across individuals for each question.

Can use “modal” responses to represent group beliefs.

(modal=majority or average response).

Can use the agreement between individuals to estimate their knowledge of the answers and the reliability of the aggregated answers.

Cultural Beliefs

Approaches for describing group beliefs:

1. SIMPLE AGGREGATION of responses across individuals (modal or average responses).
2. CULTURAL CONSENSUS MODEL: aggregates responses and estimates the culturally preferred answers to the questions, & estimates how much each individual knows the answers.

INFORMAL & FORMAL MODELS

SIMPLE AGGREGATIONS

Examples:

If responses are categories, then use mode (majority)

If responses are frequencies (or ranks), then use mean (averaged)

Reasonable:

Easy to understand and statistically sound

Best estimate of a central tendency measure for a single question

[focus is on one question at a time]

SIMPLE AGGREGATIONS

Limitation:

How can we differentiate “true” or “strong” group beliefs
from “none” or “multiple” beliefs?

In other words, how can we differentiate between a
strong cultural pattern

99% say “yes” and 1% say “no”,

from a pattern that has a majority, but no strong pattern

51% say “yes” and 49% say “no”

SIMPLE AGGREGATIONS

If the majority response is used to estimate the answer to a question, a Binomial statistical test can be used to identify responses that are significantly different from 50%:50%, e.g., can identify those questions with a strong majority.

- Limitation:

Looks at only one question at a time

(does not take advantage of information in all questions together)

CULTURAL CONSENSUS THEORY

■ INFORMAL MODELS

■ RELIABILITY:

- Average is used to estimate the answers
- Correlation between an individual's response profile and the groups' answers (item-to-total correlation)
- Square Root of reliability coefficient estimates validity

■ FACTOR OR PRINCIPAL COMPONENT ANALYSIS

- Weighted responses = answers (factor scores)
- Factor loadings = competence

RELIABILITY: AGGREGATIONS THAT USE RESPONSES TO ALL QUESTIONS

(from a series of related questions)

The reliability of an aggregation is a function of two things:

1. The number of things (# people) being combined
2. And how similar those things are

(The response patterns should be similar)

RELIABILITY

Uses MEAN to estimate answers for each question
(aggregating across people and not questions).

Can estimate individual correspondence to group beliefs:
By comparing responses of individuals
to those of the group (item-to-total correlation)

Reliability of the aggregated responses is calculated from the
number of things being combined and their inter-
correlation.

The square-root of the reliability coefficient = the
correlation between the estimated answers and the
true answers.

Weller, S.C.: Shared Knowledge, Intracultural Variation, and Knowledge Aggregation.

American Behavioral Scientist 1987, 31:178-193.

Simulated data sets were used to examine the relationship between: agreement among respondents, the number of respondents in a group, and the validity of their aggregated responses.

For a given set of answers, two different models were developed for simulating informant response data.

1. The first concerned dichotomous responses (yes/no or true/false) that were generated by a random process equivalent to coin-flipping by a computer.
2. The second model simulated interval-scale response data, also with a computer decision-making model.

(Total = 80 groups of 40 people and 40 questions in each)

Accuracy of Aggregated Responses

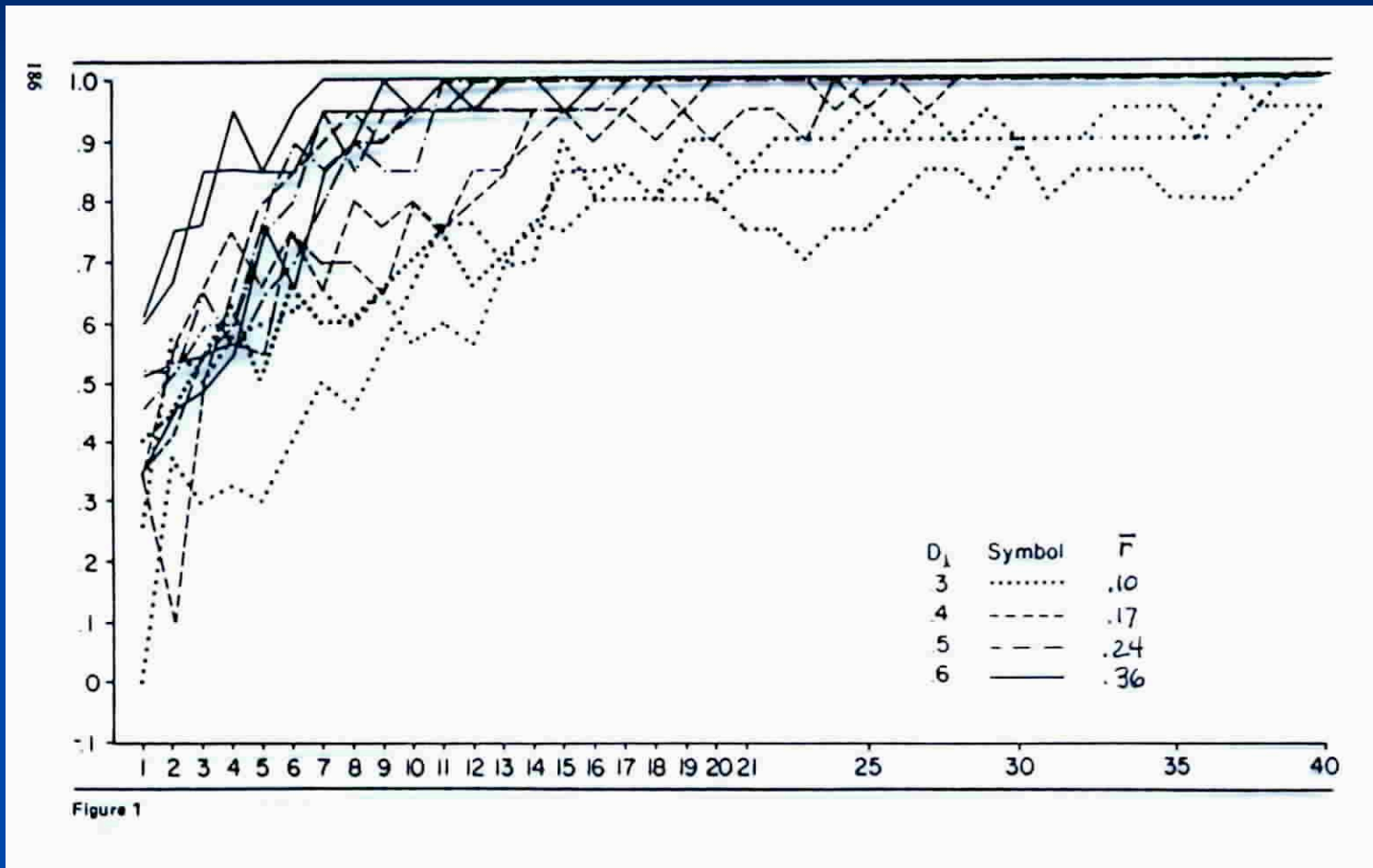
Response profiles were summed together:

with 1, then 1+2, then 1+2+3, etc. subjects.

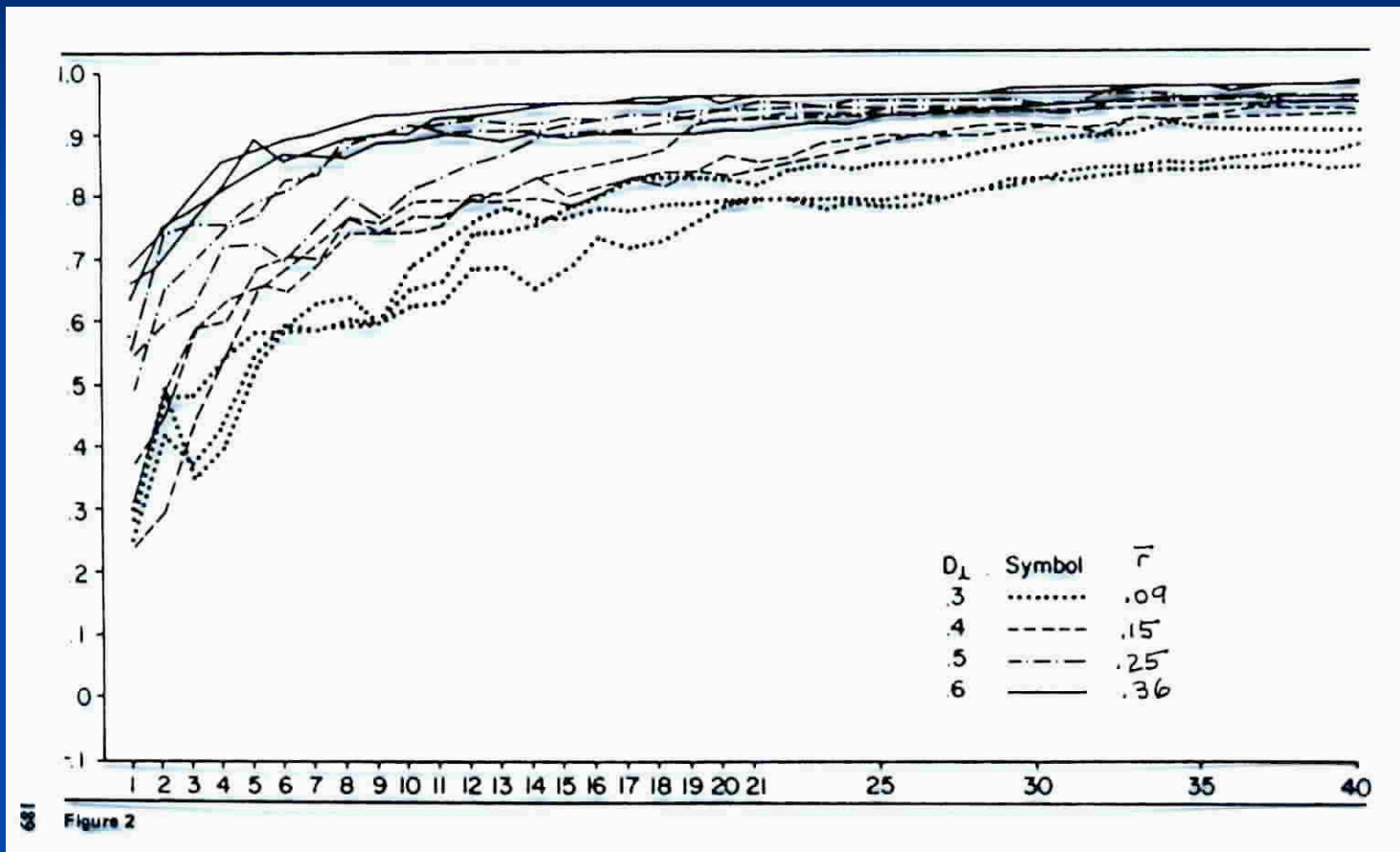
1. As agreement among informants increased, the aggregate quickly converged upon the true answers.
2. Individual cultural competency, shared cultural beliefs, and agreement among informants are directly linked:

The square-root of the average inter-informant correlation provides an estimate of the level of shared cultural beliefs among informants.

Aggregating Dichotomous Responses



Aggregating Continuous Responses



THE FIGURES SHOW HOW QUICKLY THE AGGREGATED RESPONSES CONVERGE ON THE TRUE ANSWERS AS AGREEMENT INCREASES AND AS THE SAMPLE SIZE INCREASES.

EMPIRICAL GRAPH OF RELIABILITY PREDICTIONS

It is the **SPEARMAN-BROWN PROPHECY FORMULA** APPLIED TO AGGREGATION ACROSS PEOPLE INSTEAD OF ITEMS.

RELIABILITY SUMMARY

- Answers are estimated by simple averages of responses (for interval, ranked, or dichotomous responses)
- Individual correspondence to group is estimated by the correlation between an individual's responses and the aggregated responses of the group
- Reliability coefficient estimates the degree of accuracy of the aggregated/averaged responses.
- If the average correlation between people is $> .25$,
Then 30 or fewer people are needed to estimate the answers

Romney, A.K., S.C. Weller: Predicting Informant Accuracy from Patterns of Recall Among Individuals.

Social Networks 1984, 6:59-77.

How well do informants report on patterns of communication in face-to-face groups?

Using social network data from the Bernard and Killworth studies, where both observed interaction data and participants' reports of those interactions were available.

The reliability model was used to estimate the quality of information provided by each person.

The group's aggregated reported interactions (recall data) were used as an estimate of observed interactions.

Romney, A.K., S.C. Weller: Predicting Informant Accuracy from Patterns of Recall Among Individuals.

Correlations between actual and estimated accuracy were:

- .98 for the Ham operator network,
- .92 for the fraternity members,
- .79 for the office workers, and
- .72 for the technical workers.

(There were 3 to 4 dozen people in each group.)

Findings indicated that:

1. Aggregated responses (recall data) were a good estimate of overall (observed) group behavior.
2. Comparisons of individual responses to the group aggregate, provided good estimates of the quality of information provided by each individual.

CULTURAL CONSENSUS MODEL

- The Cultural Consensus Model is a formal mathematical model for the analysis of informant responses to systematic interview questions.
- A situation is assumed in which the researcher does not know how much each informant knows about the cultural domain under consideration nor the answers to the questions.
- The model simultaneously provides an estimate of the cultural competence or knowledge of each informant and an estimate of the correct answer to each question asked of the informant.
- The model currently handles true-false, multiple-choice, and fill-in-the-blank type question formats.

Cultural Consensus Theory Assumptions

1. Common truth = Assumes a common, shared culture.
Single set of answers, shared by the informants.

Because this is an aggregation method, there must be agreement.
2. Independence of responses = Each person answers independently of every other person.
No conferring or joint responses
Consensus is measured, not created
(Procedural issue: people are interviewed separately)
3. Homogeneity of items = questions are all on a single, related topic all at the same level of difficulty.

CULTURAL CONSENSUS MODEL

The consensus model is a formal model that describes the process of asking and answering questions.

When a question is asked and the informant knows the correct answer, he/she says the correct answer.

When the informant does not know the answer he/she guesses (without bias).

CULTURAL CONSENSUS MODEL

Cultural Competency

“Competency” is differentiated from “performance”.

Competence = synonymous with knowledge, the probability (D_i) that an individual knows the correct answer to an item.

Performance = hit rate = proportion of correct answers (P_i) obtained through knowledge and guessing or chance.

$$P_i = D_i + (1 - D_i) / L \text{ (No response bias)}$$

CULTURAL CONSENSUS MODEL

Competency is estimated from the similarity in responses between all pairs of raters [Given dichotomous choices] :

a. Proportion of matches corrected for guessing/chance (L=2):

$$M_{ij}^* = 2M_{ij} - 1 = D_i D_j$$

b. Or from covariances corrected for the proportion of ones in the correct answers (π estimated at .50):

$$C_{ij}^* = \frac{C_{ij}}{\pi(1-\pi)} = D_i D_j$$

CULTURAL CONSENSUS MODEL (MATCH METHOD)

Expected proportion of matching responses between any two raters is the sum of the probabilities of the four ways that a match can occur:

1. Both raters know the correct answer and say it:

$$[D_i D_j]$$

2. One rater knows the correct answer and the other rater guesses it:

$$[D_i(1-D_j)/L \text{ or } D_j(1-D_i)/L]$$

3. Neither knows the correct answer but both guess the same response:

$$[(1-D_i)(1-D_j)/L]$$

This simplifies to:

$$E[M_{ij}] = D_i D_j + (1-D_i D_j)/L$$

Thus, proportion of matches corrected for guessing allows for est. of competency (L=2):

$$M_{ij}^* = 2M_{ij} - 1 = D_i D_j$$

CULTURAL CONSENSUS MODEL (COVARIANCE METHOD)

Similarly, the covariance between two raters can be expressed as a function of the four possible response combinations and the proportion of ones (π) in the correct answers (Z).

$$E[C_{ij}] = \pi (1 - \pi) D_i D_j$$

$$C_{ij}^* = \frac{C_{ij}}{\pi (1 - \pi)} = D_i D_j$$

Since Z is unknown, π must be approximated. Often π is assumed to be .50, since that results in the largest denominator and the most conservative estimate of competency:

$$C_{ij}^* = \frac{C_{ij}}{.25} = D_i D_j$$

CULTURAL CONSENSUS MODEL (COMPETENCY EXTRACTION)

The person-by-person agreement matrix is factored with a minimum residual factoring method to solve for the unknown values on the main diagonal (competencies).

(Comrey's MINRES provides an unbiased least-squares est.)

Since only a single factor solution is appropriate, two goodness-of-fit criteria are used:

1. Ratio of first: second eigenvalues ≥ 3
2. All competencies should be positive
($0.0 \leq \text{competencies} \leq 1.0$)

INFORMAL CULTURAL CONSENSUS (RELIABILITY ANALYSIS *ON PEOPLE*)

A set of analytic procedures and not a formal model.

(No corrections for guessing)

Reliability analysis (on people):

1. Estimates how well each person's answers correspond to the group (item-to-total correlation).
2. Estimates answers by averaging responses across people.
3. Estimates validity of the answers from the agreement between people and number of people.

INFORMAL CULTURAL CONSENSUS (FACTOR ANALYSIS *ON PEOPLE*)

Factor analysis on people

(with ML or PAF factoring method and no rotation)

Answers = wtd combination of responses

(first set of factor scores)

Competency = correlation between individual responses and first set factor scores (factor loadings)

Can accommodate:

Interval or ratio and most ordinal data

(Reliability can also handle dichotomous)

CANNOT HANDLE MULTIPLE CHOICE

Frequently Asked Questions

Development of interview materials

20+ questions, same topic, same format

“I don't know” cannot be a choice

(but can be replaced with a coin flip).

Free-lists or narrative data?

Only data with multiple questions, single responses.

(Could code and use reliability)

Rating scales?

The more anchor points, the better.

Frequently Asked Questions

- Rating scales: use with caution.
- Missing data? Cannot have any.
 - For dichotomous responses (Y/N, T/F)
just flip a coin and insert answer for person
 - For ranked data (rating scales)
insert the mean/average of person's ratings.
- Consider deleting any item and any person with excessive missing data (10%+)

Frequently Asked Questions

Doing the analysis

Software –

formal model only in ANTHROPAC & UCINET
informal [factor] analysis in SPSS, SAS, etc.

Goodness-of-fit for both models –

ratio of first to second eigenvalue ≥ 3.0
(single factor solution)

What if the model does not fit?

Low ratio or negative competence means = no fit

Compare demographic vars to competence values
to find subgroups (gender, etc)

Cultural Consensus Model (Sample Size)

Sample size determination is similar to other types of analyses; when variability is low, power is high and small samples suffice.

Here, variability is a function of agreement among raters.

3 parameters needed:

1. Average group competency level (agreement)
2. Desired confidence level for classifying answers
3. Minimum probability of items to be classified correctly

A conservative estimate may be obtained from:

1. Low group competency (.50)
2. High posterior probability for classification (.999)
3. High accuracy (.95 items correct)

<A minimum sample size would then be 29.>

Cultural Consensus Model (Sample Size)

Table 6
Minimal number of informants needed to classify a desired proportion of questions with a specified confidence level when average cultural competence is known (confidence levels of .9, .95, .99, and .999 are included).

Proportion of questions	Average level of cultural competence				
	.5	.6	.7	.8	.9
<i>.90 Confidence level</i>					
.80	9	4	4	4	4
.85	11	6	4	4	4
.90	13	6	6	4	4
.95	17	10	6	6	4
.99	25	16	10	8	4
<i>.95 Confidence level</i>					
.80	9	7	4	4	4
.85	11	7	4	4	4
.90	13	9	6	4	4
.95	17	11	6	6	4
.99	29	19	10	8	4
<i>.99 Confidence level</i>					
.80	15	10	5	4	4
.85	15	10	7	5	4
.90	21	12	7	5	4
.95	23	14	9	7	4
.99	*	20	13	8	6
<i>.999 Confidence level</i>					
.80	19	11	7	6	4
.85	21	13	8	6	4
.90	23	13	10	8	5
.95	29	17	10	8	5
.99	*	23	16	12	7

Note: *Well over 30 informants needed.

Reliability Model (Sample Size)

Again, variability is a function of agreement among raters.

2 parameters needed:

1. Average group agreement level
(average inter-person Pearson correlation)
2. Minimum validity
(correlation between estimated answers and true, unknowable answers)

A conservative estimate may be obtained from:

1. Low group agreement (.25, like competency of .50)
 2. High validity (.95 validity coefficient)
- <A minimum sample size would then be 28.>

Reliability Model (Sample Size)

SAMPLE SIZE ESTIMATES FROM INTER-RATER CONCORDANCE AND VALIDITY

TABLE 11.1
Agreement Among Individuals and Estimated Validity
of Aggregating Their Responses for Different Sample Sizes

Agreement	Validity				
	.80	.85	.90	.95	.99
.16	10	14	22	49	257
.25	5	8	13	28	148
.36	3	5	8	17	87
.49	2	3	4	10	51

Average inter-person Pearson Correlation Coefficient.

Estimated Pearson Correlation Coefficient (Square Root) between the estimated answers and the "true" answers.

More Questions

Could there be additional subgroup agreement even if the model fits?

Compare demog variables with 2nd factor loadings

(See Berges et al paper)