"Maybe" Should Be a Choice in "Yes-No" Questionnaires

Robert A. Benfer, Jr. ,[1] N. Louanna Furbee[1], and John Storer[2]

Abstract

By permitting respondents to answer "don't know" or "possibly" to queries otherwise intended to be "yes" or "no" questions, a new dimension of the interview becomes available for analysis. By coding "yes" and "no" as "yes, I have an opinion that I will express to you," and "don't know" or "possibly" as "no, I do not have an opinion that I will express to you," understanding the patterning of such responses becomes an interesting research question.

We present an example of this approach from 538 interviews, question-frames about the domain "illness," obtained in the homes of participants in an intervention program designed to reduce coronary heart disease, in central, rural Mississippi. The questionnaire was presented by four interviewers. Each participant was interviewed four times at six-month intervals, providing adequate time between interviews for reflection on the task. We use the individual differences model of multidimensional scaling to obtain weights for each consultant on each dimension of the group aggregate space. Subsequent analysis of these weights was (1) by general linear model analysis of variance and (2) examination of the pattern of adjusted means of dimension weights by risk factors and design factors.

Results were surprising. The two-dimensional aggregate space developed from opinions vs. lack of expressed opinion on individual questions was interpretable as one cluster of symptoms that implied heart disease and two other structures that were vector-like in appearance. Extremely high F-values showed a reflexive effect; the interviewer was associated with several factors including risk status of the participants, suggesting negotiation of whether or not a participant would agree to express an opinion. There may have been a reflective effect with changing patterns developing over the course of the repeated interviews. Neither dimension was associated with the health-care seeking behavior of consulting a doctor.

[1] Department of Anthropology, University of Missouri-Columbia, Columbia, MO 65211;
[2] Graduate School, Western Georgia University, Carrollton, GA 39918.

In a study of coronary heart disease (CHD) in rural Mississippi (Benfer et al., n.d.[3]), we found that multiple interviewers and multiple administrations of the same questionnaire increased the range of patterns of responses that could be usefully analyzed. Briefly, the design was an emic one in which Furbee elicited terms relating to illness and their causes, explored the domains using general pile-sorts, triadic sorting tasks, and extensive interviews based on questions that could be reasonably asked about them by interviewers. All of the interviewees were participants in a long-term intervention program whose original goal was to improve nutrition (Storer and Frate 1990) and later, to decrease coronary heart disease (Frate, Johnson, and Sharpe, 1984). Therefore, it is not surprising that many terms related to heart disease were elicited. Results from these exploratory studies were combined into a questionnaire, the Missouri Health Beliefs questionnaire (MHB). In it, all possible combinations of 19 questions for 21 terms were presented in a random order. They were asked orally, since there were many preliterate participants in the program. Table 1 and Table 2 present the question-frames and the terms included in the questionnaire. We used the Indscal model, the individual differences variant of non-metric individual multidimensional scaling (Kruskal and Wish 1990) to embed question-frames as points in two-dimensional space, which were related at least monotonically to the similarities obtained from cosines among question-frames over the 20 terms. In this, a three-way analysis, terms, questions, and individual questionnaires are permitted to vary with respect to the weight necessary to produce the aggregate, group configuration. Individuals can be associated

3 Available at http://rcp.missouri.edu/bobbenfer/index.html [To be made public soon]

Table 1: Question Frames Elicited Locally for the Domain Illness

(Terms for blanks are obtained from illness terms in Table 2) underlined words label points in Figure 1)

1. Is (are)_____ contagious?
2. Is (are) _____ an illness?
3. If you had _____, would it make you very worried?
4. Is (are) _____ a symptom?
5. Is (are) _____ serious?
6. Does (do) _____ indicate high blood pressure?
7. Is (are)_____ related to heart disease?
8. Can eating salty foods cause _____?
9. Can smoking cause _____?
10. Does (do) _____ require the immediate care of a doctor?
11. Can a person treat _____ at home without seeing a doctor at all?
12. Are nerves involved with _____?
13. Is (are)_____ a genetic problem?
14. Would home care suggest _____?
15. Does (do) _____sometimes bring along another medical problem with it?
16. Can eating fatty foods cause _____?
17. Would it alarm you to be/have _____?
18. Can a person do anything to prevent _____?
19. Does a person throw up with _____?

Table 2: Terms Elicited Locally for the Domain Illness

1. Chest pains
2. Heart running away
3. Stroke
4. Short of breath
5. Heart beat irregular
6. Loss of Appetite
7. Tired in chest
8. Sugar diabetes
9. Pneumonia
10. Stomach running off
11. Headache
12. Fever
13. Cancer
14. Aids
15. Indigestion
16. Fainting
17. Arm feeling funny
18. Arthritis
19. Sore throat
20. High blood pressure
21. Measles

with the group presentations by their weights on each dimension. We investigated the influence of the major factors of ethnicity, interviewer, administrative order, gender, age groups, location, habitat, risk status, and selected interactions by analysis of the means of the weights for each interviewee on each interview on each dimension.

Questions were of the general form: Is sugar diabetes a severe illness, where "sugar diabetes" was an illness term, and "Is _____ a severe illness" was a question-frame. We recorded the following responses.

Table 3: Original Coding of Responses

1 = Yes
2 = Maybe
2 = Sometimes
3 = No
4 = Don't know
. = Missing

Items were missing for a variety of reasons, and we will not discuss here the patterning of missing data except to note that the overall general linear model of weights from a two-dimensional Indscal of the data coded as "Missing" or "Not missing" did not quite reach the 0.05 level of significance for either dimension.

We focus on a different analysis here, one where we combined "Yes" and "No" into a new variable "Have an opinion," with "Maybe" and "Sometimes" folded into a new variable, "No opinion."

Thus, we recoded the data as follows:

Table 4: Coding of responses for different analyses

Original Code	Indscal Analysis Code
1 = Yes	1 = Yes
2 = Maybe	2 = Maybe
2 = Sometimes	2 = Sometimes
3 = No	1 = No
4 = Don't know	4= Missing

The purpose of this coding was to permit analysis of the patterning of responses where the participant in the study either expressed an opinion in response to one of the question frames and had an opinion that she or he would agree to share with an interviewer, or not. The interviewer was in the consultant's home with a pencil poised to write down the response. We wondered whether there might be differences in responses that varied with the interviewer asking the question, or whether the patterning might change over the four interviews, repeated every six months for two years. Interviewers, like the respondents, were both black and white; alll were women; three conducted multiple interviews, and one conducted the final interview for all interviewees. We were interested in learning if there was any relation between the interviewee's status of risk for coronary hearth disease and her or his care seeking behavior.

Our sample was comprised of 540 interviews of which 538 were sufficiently complete for analysis. These were analyzed using the rank variant of individual differences multidimensional scaling analysis, Indscal as implemented in program MDS (mainframe SAS, version 8.0). We used the general linear model (GLM, SPSS, version 10 for the Macintosh) in order to investigate variation in dimension weights by risk factors, race, gender, age, and in addition, degree of ruralness, and habitat, which we

thought might be related to geophagia in the study area (Vermeer and Frate 1979). We also investigated the research design effects of order of repetition of interviews and interviewer, as well as risk status for cardiovascular disease. There was a good balance for factors except for administration and interviewer. Due to an unexpected reduction in funding, it became impossible to complete the envisioned set of repeated interviews, or to complete them in a balanced design. Each participant was interviewed four times (a few fifth interviews were made but these are not included in this analysis). Although a completely balanced design could not be always followed, there was a good balance for Risk Status (271 at Risk, 267 Not at Risk), Race (269 Black and 269 White), Gender (271 males and 267 females), Rural/Urban (298 rural, 240 "urban"), whether seen a Doctor in the previous 6 months (332, yes, 206, no), and age groups (206 were less than 40 years old, 219 were 41-64, and 113 older than 64). Location was limited by the population size in the delta (42) or hils (296). Table 5 shows the number of interviews by interviewer, administration order, risk status, and race. As can be see, the four Interviewers varied considerably in the number of consultants the visited (233, 111, 109, 95). One hundred and ninety-two consultants received the first administration of the test, 192, the second, but only 88, the third, and 76, the fourth. Interviewer D administered all but one of the third presentations, and she made only a few in the others. Inteviewer 1, who made the most interviews in the first

two Administrations but only one in the third, completed 74 of the 76 interviews in Administration 4.

Table 5: Sample Sizes by Interviewer, Administration Order, Risk, and Race

Interv.	Administration Order				Risk		Race		
	1	**2**	**3**	**4**	**Yes**	**No**	**Black**	**White**	**n**
A	81	67	1	74	113	110	53	170	223
B	56	55	0	0	45	46	76	35	111
C	52	57	0	0	63	46	92	17	109
D	3	3	87	2	50	45	48	47	95
n	192	182	88	76	271	267	269	269	538

Interviewer A interviewed primarily White participants in the first administration, but interviewed equal numbers of Black and White participants in the fourth. Interviewers B and C interviwed primarily Black participants—in the case of C, 84% of her interviews were with Black participants. Thus, Administration 1 was primarily of White , and administration 3 was primarily of Black, participants. This confounding of factors because of the unbalanced design makes interpretation of results more difficult.

RESULTS

Results of the Indscal aggregate model are presented in Figure 1. A stress of 0.36 was obtained. Stress values for three-way designs are typically higher than for the simpler two-way model. Having no theoretical justification for more dimensions, we selected two for analysis.

It was a surprise to us that this two-dimensional solution would be so interpretable, given that the similarities among questions were measured entirely by whether the respondant would offer a definite opinion or refuse. The lines in Figure 1 are drawn by us to show suggested relations among the points. We see a cluster on the left in which symptoms that imply heart disease are connected by arrows. The central point, high blood pressure, is perhaps the best known harbinger of heart disease. Throwing-up is perhaps the least expected link, although nausea was thought by some to be associated with heart disease. On the right of Figure 1, we see two vectors, curved, as is typical of the representation of vector s in non-metric multidimensional scaling space. Multiple vectors may be found in these representations. In another study, we reported two vectors that traced personal and social factors towards reaching a decision as to whether to present early in a pregnancy or, instead, present late (Benfer et al. 1991), a pattern not observable in a classical multiple regression representation (McKinney 1987, Fisher et al. 1991).
In the present instance, we interpret the shorter vector as representing infectious diseases, with recognizable symptoms that are preventable and can be treated at home. The second, longer vector is one that orders progressively more serious terms, possibly terms for chronic or acute illnesses. More speculatively, one can also trace a path from Heart Disease, to the vector of Alarm, Worried, Doctor's Care, Serious, Illness, and Illness which Brings Other Medical Problems.

If we revisualize this suggested internal structure of cluster and vectors as the two vectors produced by the multidimensional scaling, we can interpret Dimension 1 as contrasting heart disease with other illnesses. In Dimension 2, we see that Heart Disease, Alarm, and Worried define one end of the vector whereas Contagious, Genetics, Symptom and Brings other Medical Problems are weighted most strongly in the opposite direction. We interpret this dimension as measuring the strength of expressed knowledge of illnesses of the heart. In our experience, we have found it not unusual for the second dimension to somewhat mirror the first in Indscal, since our work has always been within a single domain, where very many distinct, uncorrelated dimensions would not be expected.

Despite our previous experience in finding interpretable vectors, vectors are not the most common structures unfolded from these kinds of data (Shepard 1980, Furbee and Benfer 1983). Nonetheless, with the inclusion of interaction terms, the vectors can be

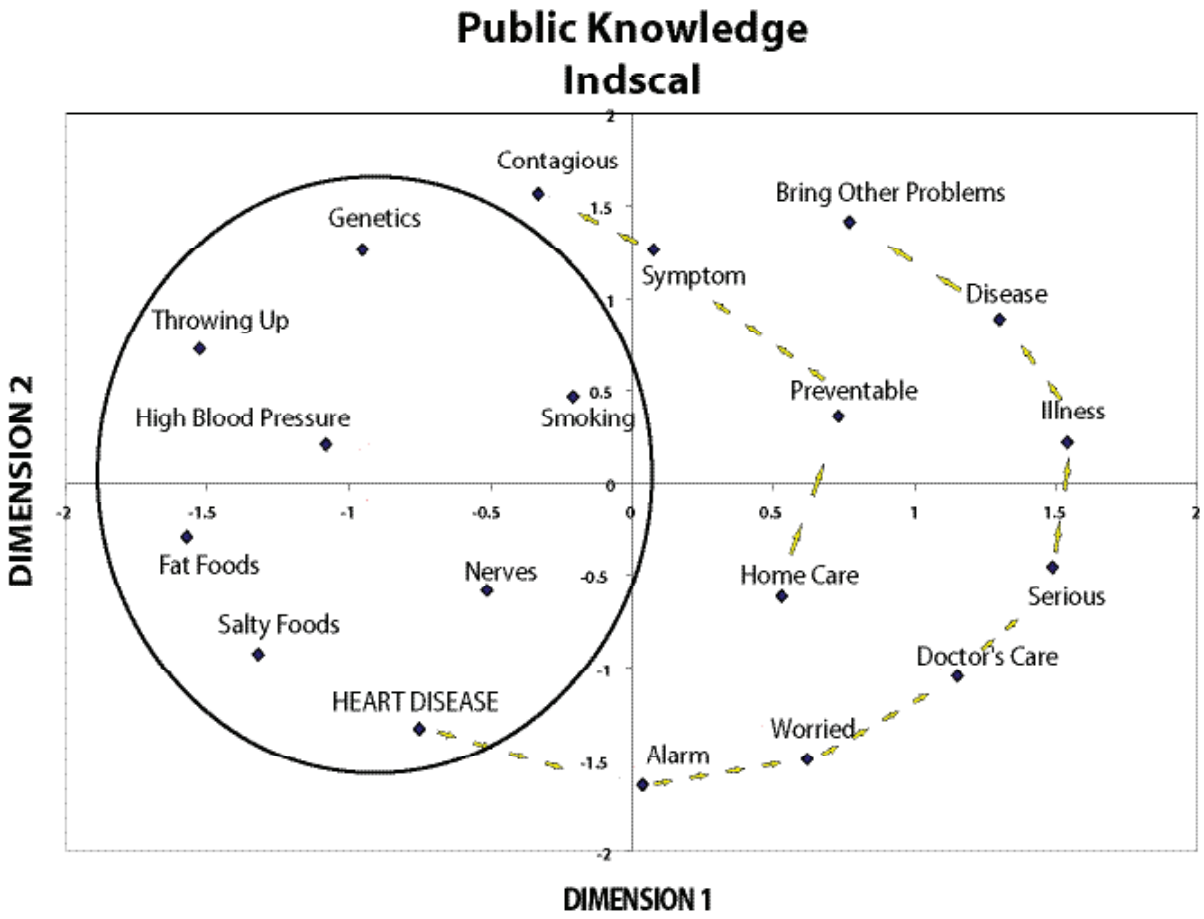

Figure 1 Two-Dimensional Indscal Aggregate Solution

evaluated by the patterning of personal weights of each participant. These personal weights are the weights that would transform the group representation of Figure 1 to one that more closely represents the cognitive map of the individual. Means that vary significantly by risk or design factors can be further analyzed for the patterns of their means across the factors.

Table 5 presents analysis of variance results obtained from the general linear model. The GLM combines multiple regression with ANOVA, and is used here to evaluate variation dimension weights by individuals grouped by the risk and the design control factors. The dimension weights did not exhibit significant skewness or kurtosis and thus are suitable for this linear model. As can be seen, the weights for individuals,

when grouped by risk and control factors, vary considerably along both dimensions. Dimension 1, which we interpreted as contrasting heart disease with other illnesses varies most strongly, reflexively, by interviewer. We noted some problems in inter-interviewer reliability in the pilot study phase of the project and instituted more training for the three interviewers. Unfortunately, due to the unexpected budget cut, one interviewer conducted most of the third set of interviewers, and a new person, who was brought in for the fourth set. This fourth interviewer was not trained except for accompanying Interviewer C when the latter conducted one of her third set of interviewes.

Figure 2 presents the least squares means of Dimension 1 weights; these are means that have been adjusted to estimate those that would have been obtained if the risk factors had not been intercorrelated, or, alternatively, if scores had been average on all the other factors. Thus, they offer statistical control for a design that was not completely balanced. Figure 2 shows that there is a strong interaction between Interviewer and status for Risk for CHD for Dimension 1. For the first three interviewers, trained by Furbee, interviewees not at risk consistently scored higher than those at risk for the dimension, as if they emphasized in their willingness to respond that heart disease is not so very different from other illnesses. With interviewer D, we find the reverse, suggesting that somehow in her interaction, during the approximately two-hour interview, she elicited the opposite pattern, that heart disease was a more serious illness.

Table 5: F-values: Analysis of Variance from GLM: Indscal Dimension Scores by Risk and Design Factors (Mean Squares from Type III SS)

		DIMENSIONS	
FACTORS	df	**1**	**2**
Administration	3	**18.07******	**7.71*****
Risk	1	0.10	0.75
Race	1	0.08	0.01
Interviewer	3	**17.05******	**6.95*******
Gender	1	0.83	0.86
Delta/Upland	1	0.19	0.25
Rural/Urban	1	0.32	0.62
Seen physician 6 month	1	0.61	0.48
Age group	2	1.91	1.81
Admin*Race	3	**2.89***	1.92
Admin*Risk	3	**21.25******	**8.06******
Race*Risk	1	1.58	0.73
Interv*Risk	3	**22.87******	**9.12******
Gender*Risk	1	0.00	0.21
Delta/Upland*Risk	1	0.23	0.32
Rural/Urban*Risk	1	0.64	0.60
Doctor 6 month*Risk	1	0.00	0.19
Age group*Risk	2	0.62	0.41
Race * Interviewer	3	**3.18***	1.97
Race*Interviewer*Risk	3	0.32	0.44

*P< .05, ** P < .01, **** P < 0.001,***** P < .0001

Figure 3 shows the least squares adjusted means of Dimension 1 organized by Administration order. Recall that interviewer D conducted most of the third interviews., while interviewer A conducted most of the last. Figure 3 may be displaying her reflexive, interviewer effect. However, since this is the third interview, the effect might also be one due to greater time for reflection on the questions by the interviewees. In general the same patterns are repeated in Dimension 2 although the effect of Interviewer by Risk status, while less strong than for Dimension 1, is relatively stronger than are the other factors for Dimension 2.

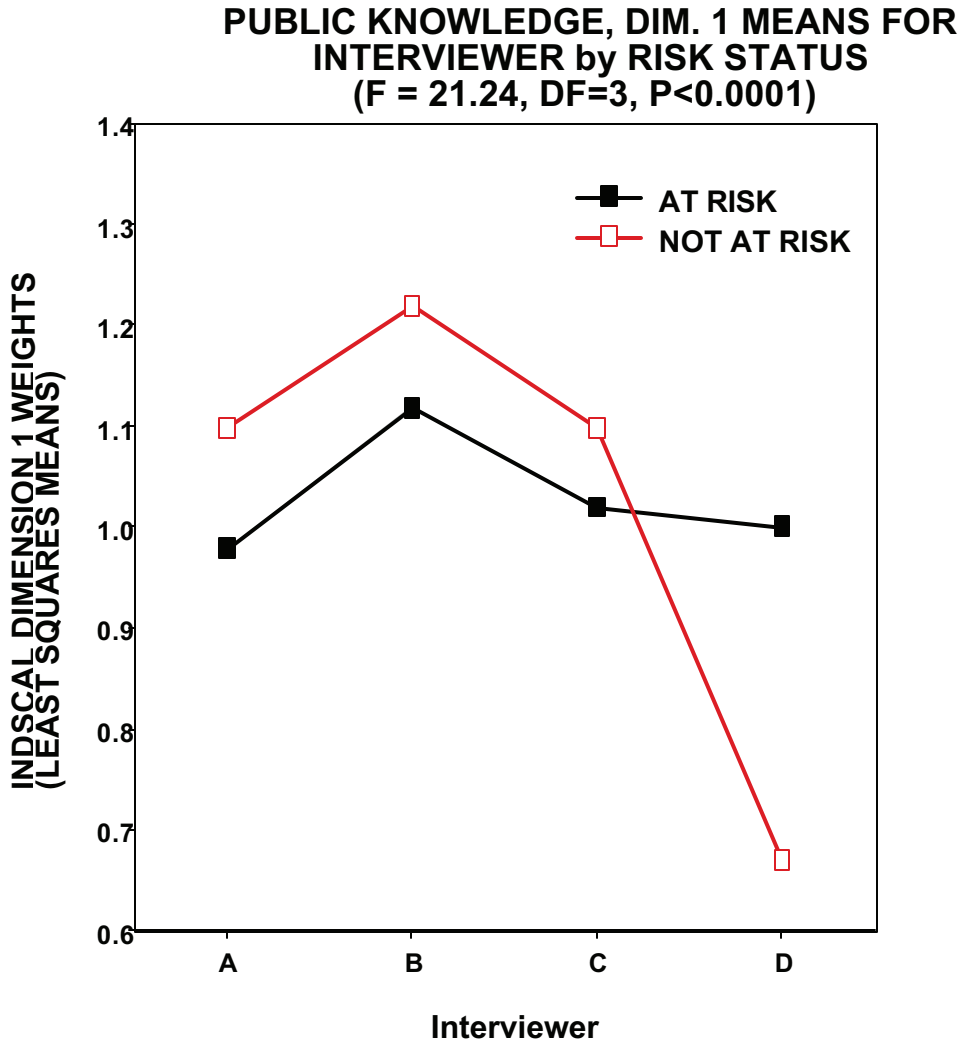

Figure 2 Adjusted Dimension 1 Weights by Interviewer

In order to try to understand the interview effect better, GLM analyses were made separately for each interviewer, for Dimension 2. Recall that we interpreted this dimension as measuring strength of response to heart diseases. For this dimension, questionnaires conducted by Interviewer A showed Race as the strongest effect (F=

24.04 with 1 df, P < 0.0001). Next in importance for As' interviewees was the interaction of Risk Status with whether the person had Seen a Doctor in last six months (F = 6.38 with 1 df, P=0.01), followed by less strong influences of the Administration Order by Race interaction (F = 3.26 with 2 df, P =0.04) and Rural/Urban location (F =3.69 with 1 df, P = 0.06). Interviewer B showed only one significant factor, that of Race (F = 4.68 with 1 df, P = 0.03). Interviewer Cs' data showed a strong change in response over the two administrations that she made of the questionnaire (F = 10.06 with 1 df, P = 0.0002). Interviewer D, who conducted the last and smallest set of interviews, also showed a Race effect (F = 6.88 with 1 df, P = 0.01), but no other effect was significant. Contrasting these results to the pooled analysis reported in Table 5, we can see that Interviewer A, whom we considered to be the best interviewer at the time of the training, probably contributed the most to the pattern of significant results. The administration order effect may have primarily been one elicited by Interviewer C. Interviewer D, whom we did not train, elicited the strongest responses with respect to risk status of any interviewer.

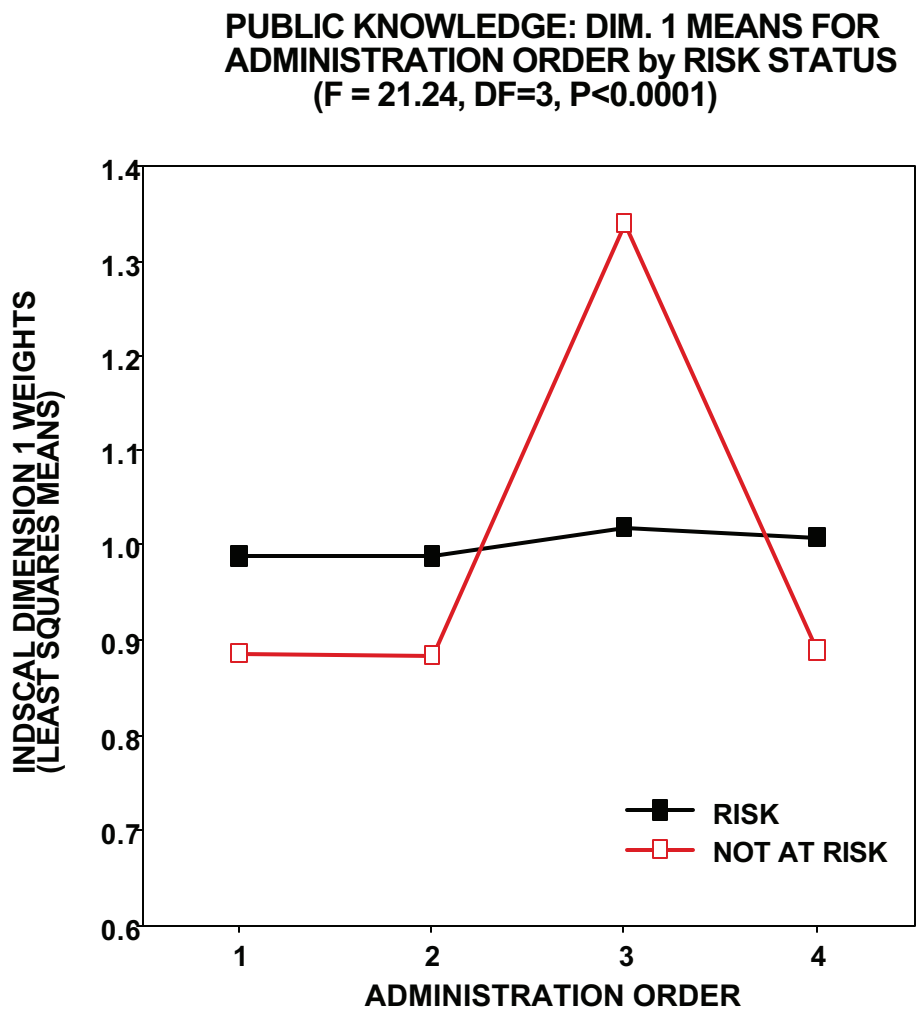

Figure 3 Adjusted Dimension 1 Weights by Administration Order

Conclusions

We first note the reflexivity demonstrated by our four interviewers with the participants in the study. In the more traditional study, reported elsewhere, where "Yes" and "No" were distinguished and coded as in Table 3, "Race," an index of ethnicity, was a very strong factor; it is not so strong here where we are only measuring willingness or knowledge enough to respond with an opinion. Although we did find a

barely significant effect in the Race by Interviewer interaction term for Dimension 1 ($F = 3.18$ with 3 df, $P < 0.05$), its low magnitude suggests that reflexivity between black and white interviewers and black and white participants is probably not a primary source of variation. However, we did not analyze the data separately for the race of interviewer by race; the interaction terms presented in Tables 1 and 2 were across all four interviewers and variation doubtless includes both personal characteristics independent of race as well as race.

The reasons for the interpretable pattern in Figure 1 are probably diverse. Some participants may not have wanted to answer, for example, consider whether one would want to answer "Yes," to the Question Frame "Is Tired in the chest a Serious illness?" One can imagine that a White participant might not have wanted to express an opinion if he or she thought it more commonly used by Blacks, or perhaps the participant felt that the choice of words, if responded to, would label him or her as lower in class. Of course, some may have not responded because they were not certain; slight paralinguistic signals and body movements by the interviewer might also have encouraged respondents to produce the "right" answer. Clearly, further research is needed.

In the present study, risk status varied significantly, by whether or not a consultant offered any opinion. This variation was probably primarily due to reflexive effects of two of the four interviewers (C and D). That we thought C the least acceptable interviewer of the initial three and D was not trained by us at all is provocative. We conclude that rather than viewing inter-interviewer variation as objectionable, analysis of Indscal dimension weights converts this potential problem into a useful new source of insights. The patterning that can be produced by a single investigator with a single

replication may not be as full as one in which the reflexive effects of multiple interviewers is present, and the reflective possibilities for the consultants between questionnaires may combine with this reflexivity to produce a richer understanding of a domain.

Multiple presentations of the same questionnaires (Benfer and Furbee 1989) suggest the possibility that a consultant might be best interviewed after having reflected on the particular domain presented as a questionnaire in previous repetitions. We argued that the informant would have time to begin to develop a more consistent model of the task, which is novel when first presented. However, because of our inability to balance interviewer and replication here, the two are partially conflated, and we are unable to accurately estimate separate effects for reflectivity and reflexivity.

Recommendations

These results suggest that it may be valuable to collect questionnaire data in which "Maybe" is a permitted response. Contrasting having an opinion to express, with not expressing an opinion provides additional dimensions for analysis. Multiple presentations of this kind of questionnaire, or other kinds, may capture additional dimensions developed by the consultant while reflecting about previous interviews. Multiple interviewers may elicit different patterns of meaningful responses reflexively. Written instruments might permit studying the reflective effect independent of the reflexive one, since presumably reflexivity would be reduced. That strategy was not possible in this case where over one-fourth of participants were not literate.

Permitting a consultant to deny holding an opinion instead of insisting on his or her expressing knowledge about a particular question may open new avenues of research, opportunities for which may be presently lying quiescent in questionnaire data

sets where that response has been recorded. In any case, we advise that when preparing a survey instrument where the intent is to elicit yes/no responses, keep in mind the possible reflexivity of oral interviews and the additional information provided by repeated interviews of the same consultants by multiple interviewers.

References Cited

Benfer, Robert A., and Louanna Furbee 1989. Procrustes Analysis of Individual Configurations: Patterns and Axes Similarities. *Journal of Quantitative Anthropology* 1:65-80.

Benfer, Robert A., Edward Brent, and Louanna Furbee 1991 *Expert Systems*. Sage Press, Newbury Park, CA.

Benfer, Robert A., Louanna Furbee, John Storer, Julie F. Farnum, and Dennis F. Frate n.d. Health Beliefs as Predictors of Health Seeking Behavior: Coronary Heart Disease in Rural Mississippi. (available SOON at http://rcp.missouri.edu/bobbenfer/index.html)

Furbee, Louanna, and Robert A. Benfer 1983 Cognitive and Geographic Maps: Study of Individual Variation among Tojolabal Mayans, *American Anthropologist* 85:305-334.

Kruskal James B., and M. Wish (1990). *Multidimensional Scaling* (15th reprinting). Sage Series in Quantitative Applications in the Social Sciences, No. 11. Newbury Park, CA.

McKinny, Gregg, 1987. A Model for Predicting the Initial Presentation of Pregnant Women for Prenatal Care. Master's Thesis, University of Missouri-Columbia.

Fisher, Mary Jo, Bernard Ewigman, James Campbell, Robert A. Benfer, Louanna Furbee, and Steven Zweig 1991. Cognitive Factors Influencing Women to Seek Care during Pregnancy. *Family Medicine* 23:443-426 (with Editorial).

Shepard, R. 1980. Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, 210: 390-398.

Storer, John H., and Dennis A. Frate 1990. Hunger, poverty, and malnutrition in rural Mississippi: Developing culturally sensitive nutritional interventions. *Human Services in the Rural Environment* 14:25-30.

D. E. Vermeer, and Dennis A. Frate 1979. Geophagia in Rural Mississippi: Environmental and Cultural Contexts and Nutritional Implications. *American Journal of Clinical Nutrition* 32:2129-2135.