Sociology 357/8 Pamela Oliver

STRUCTURED QUESTIONNAIRE EXERCISE

DUE DATES TO CHECK IN CLASS

** First draft of questions due. (Whole group submits one, or if different people have different ideas, submit all sheets as one set fastened together.) Do this as soon as you can, so you can get my suggestions and get going.

** Completed code sheets due for computer analysis. (Remember to attach copy of questionnaire.) Whole group uses exactly the same questions and exactly the same code sheet formats, and submits all data together as a set. Put all group members' names and phone numbers on the code sheets, in case we have problems.

** Final report due.

Your task for this exercise is to develop and pretest a set of questions which could be used to test a simple bivariate hypothesis. Your pretest will involve administering the questions to a minimum of 20 people (a minimum of 10 per group member), checking the inter-item correlations to be sure it is appropriate to sum them in a composite index, and testing your bivariate hypothesis with your index. You will collect data from a convenience sample, but try to get diversity in what you are studying. To make it feasible to do this in a short time, your topic should be something it is meaningful to study on the population of university students or some other population to which you have very easy access within the time frame of this project.

About Teams

You are encouraged to do this exercise in teams of 2 or 3 people. (The only disadvantage of a team is scheduling, but this will be outweighed by the advantages unless you have a very complex schedule. Developing good questions is a task better done by several heads than one; teams can collect more data and can divide the labor in coding and preparing tables. Team members will normally write a group report plus an individual evaluation, and this is what I strongly prefer. If time schedules or personality conflicts make preparing a group report too difficult, some or all of the report may be written separately by each group member (or subsets of the group). If individual reports (or sections of reports) are submitted, you must include a statement indicating who worked together or consulted how much on what sections.

A team of 4-5 people may write a questionnaire together and then break into teams of 2-3 to write. You may also form super-teams of 8-10 people that develop the same questionnaire (so you have more cases in your analysis), but in this case, you should have two different indices (developed by different sub-teams) that it makes some sense to relate to each other. Writing teams should be 2-3. We can discuss these options and their mechanics more in class. More cases are better for analysis, but we don't want the coding & data entry task to get out of hand.

Questionnaire Development

NOTE: As stated on the syllabus, each team should submit a rough draft of your questions to me for comments as soon as possible. Submit all the ideas if you disagree. In general, if you agree on what concept to measure but disagree about which are the best questions, you can usually solve the problem by including all the possible questions in your questionnaire and let the statistical analysis tell you which are the good ones. I will be available during class periods and for extended office hours to give groups individual assistance on this.

ANOTHER NOTE: The instructions below assume that your multiple-item index is dependent and everything else is independent. It will actually work out fine if your multiple-item index is independent and you have a simple attitude or behavior for your dependent variable; you would still include a few additional independent background factors as possible control variables. This would require minor changes in the organization of the write-up. See me for details if you get into this.

1. Dependent variable. Pick a relatively general attitude, belief, or behavioral pattern, one which could obviously have a wide variety of possible measures. This concept must be at least *ordinal*: that is, there must be some single dimension which you can have more or less of. Almost all concepts fulfill this criterion, so it is not likely to be a problem for you, but ask if you are not sure. Your task is to develop a variety of closed-ended questions which can be summed to form an index which measures this concept with less error than the individual items would have. About 90% of your effort in writing questions should go into these multiple measures of the dependent variable. Additionally, you will write one general open-ended question to capture the person's ordinal opinion in his or her own words; we will use this question as a validity check. You may optionally ask an additional open-ended question to elicit additional information like the "reasons" in Schuman's article (pp. 267 ff in the Golden reader).

1.1. *Closed-ended questions*. There should be 6-10 closed-ended questions to measure different aspects of the same dependent variable. One of them should be a straightforward general question to capture the main idea of what you are interested in. The others should ask about different dimensions or themes relevant to your attitude. Unless this is impossible, some of these items should be worded positively so that agreeing means a person is at the high end of your concept, while others should be worded negatively so that disagreeing means a person is at the high end of your concept. See me if you don't think you can do this.

The response categories for the questions should be ranked to express more or less of the attitude you are measuring. They should be printed in rank order on the questionnaire. You should generally have four to seven response categories for each question, although there are exceptions.

TO CHECK YOURSELF: You will end up wanting to add up these items, so it has to make sense to do this. All the different questions measuring the dependent variable should give you a score of high, medium, or low on the same conceptual variable; + items will yield a high score for agreeing, while - items will yield a high score for disagreeing. If you have difficulty deciding whether a particular item should be a + or a -, it probably has

a problem. If you have difficulty with most of the items, either they or your concept (or your understanding) have more serious problems, and you should see me.

1.2. *Open-ended question*. Include in your questionnaire one general and straightforward open-ended question that asks people to use their own words to tell you what they think about the issue that is the subject of the dependent variable. You will categorize according to people's general attitude; this will provide useful information for evaluating your closed-ended questions.

2. Independent variables. Select a few (2-5) independent variables for which it is easy to write simple closed-ended questions. Prepare these simple questions and include them in your questionnaire. These should be variables that you believe will be related to the dependent variable **and** which **vary** in the UW student population (or the population you will sample).

Choose one independent variable which you really believe affects your dependent variable; this variable will be the one you want to use for your hypothesis. It may be a basic background variable, or may itself be an attitude or behavior which you can measure with one or two items. The other independent variables are background factors you suspect will make a difference in your dependent variable. These are to give you the possibility of having additional information if you need it.

These independent variables can be nominal, ordinal, or interval. We will make sure you get back statistics appropriate for your level of measurement. Everything we will be doing comes from the "bivariate association" part of statistics.

3. Optional Additional Variables. You may choose to create more than one index, either for a different dependent variable, or for a more complex independent variable. These are created in the same way as the instructions say for the basic dependent variable. There is an example of this in my sample paper.

4. Refine Questions. Following the principles described in the text and in class, refine your questions to make them as good as you can, both in their content and in the physical structure of your questionnaire or interview schedule. All questions should meet formal criteria such as: a) Unbiased; if biased items are used, they should be balanced. b) Clear, unambiguous, single-barrelled, grammatical. c) Closed-ended categories are exhaustive, mutually exclusive, reasonable in range and precision, and fit with the stem of the question. d) Legible, logical physical presentation.

Data Collection (Teams do this part together.)

1. Make copies of your team's questionnaire. Each team must collect data from a minimum of 20 people, and each person must do a minimum of 10. (i.e. 1 and 2 person teams must do a minimum of 20, a 3 person team must do a minimum of 30, and a 4 person team must do a minimum of 40). It is actually not much harder to do more, but a total team sample size of 50 is plenty for the statistics and as much data-entry as we want to handle.

2. Do convenience sampling, but purposively try to get as much variation as possible *on the variables you are studying*. That is, try to get people you expect to have different opinions on your dependent variables and to differ on your independent variables. Do not just ask your friends, and do not get whole groups of people to fill them out together, as any chance for people to talk to each other while writing their answers is likely to produce error and bias.

3. As you collect the data, record any information that might be relevant to understanding people's answers, or that might alert you to problems with the questions.

a. You may ask the questions orally or let people write their own answers; just say which you did.

b. Ask people to tell you or to write in the margins if there is anything they find offensive, difficult to answer, unclear, etc.

c. Watch and listen to the respondents for signs of difficulty or confusion, irritation at questions, hesitation in answering an item, giving the form back to you blank, changing answers, laughing, explaining the answer, etc. Take notes about these observations on the form itself, or on a separate paper you later staple to the questionnaires.

4. Put a unique number on each questionnaire, right on the actual paper used for data collection. USE THESE SAME NUMBERS IN YOUR CODE SHEET, BELOW. Note that partners' questionnaires must all have different numbers (you cannot have two 9's, for example). One way to do this is to use three-digit numbers, where the first digit indicates the team member. In this system, 104 is partner #1's 4th questionnaire, 203 is partner #2's 3rd questionnaire, etc. It is useful to use some numbering system that makes it easy to know who collected the data. Please use numbers not letters because that makes our data entry easier.

Data Organization

1. Follow the example and instructions in "Comments on Coding" for preparing your code sheet for computer analysis. Teams may code their data together, or each individually, but <u>all</u> the team's data must be submitted together as one package, and partners' data must end up in a compatible form to be analyzed together on the computer. THE WHOLE TEAM MUST USE THE SAME FORMAT FOR THE SUMMARY SHEET. That is, the variables must be listed in <u>the same</u> order across the page. The sheet must be legible: it must be written in dark pencil or pen, the letters and numbers must be large enough to read, and it must be neat enough for someone to translate it without your being there.

If you have access to a spreadsheet program, do your data entry in the spreadsheet, export it as a text file, and email the file to me or an address I give you. I will show you in class how to “export.” THE WHOLE TEAM’S DATA MUST BE TOGETHER IN ONE FILE! The email message should give the names of the team members! (Some of you have really weird email names.)

2. Attach a copy of the questionnaire, with the variable name written next to each question. Mark the code sheet and questionnaire both to show dependent variable questions, independent variables, control variables (if any), open-ended question. If you want to request control tables, or are doing something unusual, be sure an explanation is

attached. If you send us the data in an electronic file, give us this information on a separate piece of paper. Put the names, telephone numbers, and email addresses of ALL team members on the code sheets or information page, and if you send an email file, put this information in the email message, too. Tell us who should be the "primary contact" if there is a problem.

3. Go over the questionnaires or summary sheets, discussing them and writing notes on them about problems or inconsistencies which can be seen within one person's answers. If you have the questionnaires, you can write these notes on them, or on a separate sheet of paper, whichever seems better. You can often get a good informal feel for how your closed ended questions are working by reading over the questionnaires and seeing whether people's answers on the closed ended questions seem to make sense in light of their open ended answer. Write notes on questionnaires where the pattern of answers seems odd, and keep track of this information for later use.

4. Check each other's work and be sure everyone is doing things the same way. This is extremely important. If one team member deviates from the rest, everyone's data will be garbage, no matter who is "right." If in doubt about how to follow the instructions, at least make sure that the whole team goes the same way. If you are worried that you and your partner did things differently, it will save everyone a lot of work if you at least tell us that you are worried so we can check (but it is better to talk to your partner).

Computer Analysis of Data (Done for team together)

We will have someone else type your data into the computer and generate tables for you. You will turn in your code sheets and will receive back computer printouts which give: a) univariate frequency distributions for all variables; b) correlations among all the measures of the dependent variable; c) reliability analysis for your proposed index; and d) appropriate bivariate statistics (correlation or difference of means) for relation between your index and each independent variable.

SUBMIT YOUR CODE SHEETS ON (OR BEFORE) THE DATE ANNOUNCED.

Preliminary Data Analysis

1. You will get list of your data as it is stored in the computer. Check this list against what you wrote on the code sheet to be sure there are no errors. It is also worth checking the computer list back against your original questionnaires. It is rather easy for me to correct your computer file and generate correct tables for your group. If you find that there is a mistake in your data, you will probably save time in the long run getting it fixed rather than in trying to cover it up. Data with errors is usually extremely difficult to interpret.

2. Frequency distributions. Once you know your data are ok, look at your frequencies. There are several useful things to inspect the frequencies for. A) Out-of-range values, which point to errors in coding or data entry. B) Problems of low variability. If 80% or more of the cases are in one category of a variable, the variability of that variable is so low that statistical results with that variable are extremely problematic with small samples. Results from such variables cannot be trusted, and they are often simply discarded. C) Lesser problems of low variability, where 60% or more of the cases are in

one category, or where 80% or more of the cases are in two adjacent categories (of a variable with more than three categories). These may just indicate a skew in the population, but might also indicate a biased sample or a biased question. D) Comparisons across the different questions measuring the dependent variable, to see which items elicit the most favorable responses, and which the least favorable. E) Information about the distribution of attitudes in the population. Because your samples are non-random, you have to interpret these results very cautiously, but it is still interesting to find out what people said.

3. Bivariate relations between your dependent variable questions and index and your general closed-ended dependent variable question or your coded open-ended question. If either your general question or your coded open-ended question is good, this will give you a quick check on the validity of the other questions.

4. Correlations and reliability analysis among the questions measuring the dependent variable. If different questions are all measuring the same underlying concept, all correlations should be positive and moderate to large (above .5 or .7). The reliability analysis should show a coefficient alpha that is pretty high (.9 is ideal, and we hope for above .7) and should show that removing any item from the scale lowers the alpha. We can live with weak correlations. However, significant negative correlations point to problems we have to deal with.

If you don't have the ideal, I will go over your data with you and help you figure out what to do. This is usually too complicated for beginners to figure out themselves. Patterns to check for: a) One or two questions account for all the negative correlations and most of the small positive ones; these same questions have low item-total correlations and may even show an alpha that goes UP when they are removed. We re-compute the index without these questions. We conclude that the rest of the questions provide an adequate measure of the dependent variable. b) There are subsets of questions which have moderate or strong positive correlations with each other but negative or weak correlations with the others. This pattern can make the whole reliability analysis look bad. This means the questions seem to be tapping two different conceptual variables. You try to see which set most captures what you had in mind and use those for your index. You will usually want to re-run the reliability analysis for each set separately. Occasionally, you decide that both sets are interesting, and create two indices, one for each. c) Most correlations are moderately or strongly positive, there are no negative correlations and only a light scattering of weak positive ones. This probably means there's no particular trouble at all, and the reliability analysis will look good. d) Most correlations in the table are close to zero, and negative ones are scattered across different questions. This means that none of the items are properly related to any of the others, that there really is no single concept that your questions measure. The reliability analysis will show a low coefficient alpha. This is the worst situation to be in, but it is rare. I try to give you enough help in writing the questions to avoid this. Most often, this turns out to be due to mistakes in coding the data, usually when partners code their data separately and turn out to be doing it differently. If you have data like this and you have checked for and ruled out coding errors, you definitely need to see me. e) There are a lot of strong negative correlations for one or two

variables. This usually means that you have forgotten to "reverse score" a question, or that subjects read the opposite meaning into a question than you intended. See me.

WRITTEN REPORT

PLEASE FOLLOW THIS FORMAT EXACTLY. This is based on Chapter 18 of the Singleton book, but includes some specifics for this class.

About Truthfulness. Science depends on researchers telling the truth about what really happened in their research, *not* what they wish had happened. At the same time, students worry that they will be graded down if they tell the truth. So, for each question, I insist that you tell the truth about what really happened in the research, but then follow it with an opportunity to explain what you now think you should have done. If there was a mistake and your self-criticism gives a correct statement about what you should have done, you will receive full credit as if you had done things right in the first place.

Important Stylistic Note #1. Give each question a short phrase that describes its content. This is usually related to the name, but is often somewhat longer, as many questions cannot be captured in eight letters. Use these descriptive phrases as labels in all tables and in writing about the questions, *not* question numbers or cryptic computer names. Be consistent, so that the reader can easily work between the tables, the text, and the questionnaire. That is, call the same variable the same thing everywhere.

Important Stylistic Note #2. Tables may be embedded in the text or prepared on separate pages which can be interleaved with the text or attached together at the end. Each table should have a number and a title; everything should have meaningful labels. Discussions in the text should refer to tables by number and should use the same variable names as appear in the tables. Cut-up computer printouts are *not* adequate tables. Part of writing a statistical report is arranging the results into readable tables. It is easiest to prepare tables in a spreadsheet or using the tables option in a word processor. If you do not know how to do this, it is faster to do them by hand, and I will accept this.

OUTLINE

Title page. Title of report, author(s), date. Put partner's name in parentheses at the bottom of the page if you worked with someone but wrote reports separately. If you are part of a large “super team,” the team should agree on a team name and everyone put it in parentheses, under the name(s) of the paper author.

Abstract. Write one paragraph which summarizes your hypothesis, data collection procedures, and findings. You may include this on the title page if you wish.

Body of paper.

I. **Introduction**. Write a paragraph stating your *bivariate* hypothesis and why it is worth researching. For this assignment, you will usually want to talk about why your dependent variable is interesting to you. (Note: Citations to readings are not needed, but go here if something you read went into your thinking on this project.)

II. **Methods of research**. (Note: To aid grading, number each section of this discussion as it is numbered here.)

A. **Sampling**. 1) Describe your sampling procedures including when, where, and how you selected your subjects. Although rigorous sampling is not required, try to think analytically about the coverage and biases of your sample. Describe the kinds of people you got into your sample, given your procedures. Be sure to discuss any differences among team members in this. 2) Tell the reader in a couple of sentences what to think about the external validity of this paper. We know you do not have a probability sample, so strictly speaking, your external validity is low. But less strictly speaking, do you feel that you probably have a good representation of the population of interest, or do you feel there are clear biases? Explain a little. 3) Evaluation: why you think your sampling was good, given your resources and limitations, or what you now believe should have been done differently. Please note, this evaluation is in terms of what was actually possible in this assignment, and is <u>not</u> about the standards you believe professionals should adhere to.

B. **Independent Variables**. 1) Most of these variables are simple and unproblematic. Just list them, noting anything that would not be obvious. (Example: the categories of sex are obvious, but the categories of religion are not, so you should say which you used.) If you have a more complicated independent variable, state the question(s) you used to measure it, and anything you did in the way of recoding or forming an index. See instructions for dependent variable for information about how to write about an index. 2) Either assure me that there were no problems in these, or describe any problems that turned up. 3) Briefly evaluate your measures, stating any changes that should be made. (Note: relation to dependent variable goes elsewhere, as does discussion of wishing you had included other variables. This is just evaluation of the measures of the independent variables.) 4) Discuss whether there is adequate variability in the independent variables for analysis.

C. **Dependent variable: closed-ended questions**. 1) Discuss a bit how you defined and revised your concept, and what thinking led into the specific questions you used to measure it; briefly list the items or refer to the questionnaire copy. 2) Briefly explain what it means to be on the high end (have a big number) in your concept, and what it means to be on the low end (small number). Referring to the labeled questionnaire copy in your appendix, tell me which closed-ended questions are reverse-scored and which are not. Explain what a high score and a low score mean for one of your close-ended items that is not reverse scored, and one that is. (That is, select one question that is reverse scored and one that is not, and answer this for each of those two questions specifically.) 3) Based on your own second thoughts or the comments of your subjects, evaluate the closed-ended items in terms of their clarity and lack of bias. Are they OK, or should some be revised?

D. **Dependent variable: open-ended question(s)**. 1) Describe the coding rules (operationalization) you used for grouping open-ended responses. 2) Evaluate the question and your coding: Did the question seem to "work" in giving you an overall idea of the person's opinion on the subject? Were you able reliably to group subjects into two to five categories, or were a lot of the answers ambiguous? 3) Did the content or reasons in people's answers reflect the ideas you had in thinking up your closed-ended questions? Or, did the content suggest themes or issues you had not considered?

III. **Validity of Measure of Dependent Variable**

(NOTE: If your index is for your independent variable instead of your dependent variable, just make the appropriate adjustments.)

A. Check for frequency distribution problems. (Refer to the tables in the appendix.) 1) Discuss all variables that appear to have problems with low variation, explaining what the problem is. 2) If none of your variables has a problem, summarize the variability in the variable with the *lowest* variability.

B. Summarize the results of your reliability analysis and table of correlations. (Refer to the tables in the appendix.) If some items were dropped, list them and say why. Then, for the items that remain in your index, describe the range of correlations and tell what the alpha is for the scale. Briefly say how good the index seems to be.

C. Prepare a table for the difference of means table for the index by category of the open-ended question (or general closed-ended question if the open-ended had problems). You should recopy this table yourself and properly label it, do not just cut out the printout. Do these two measures appear to be consistent? Discuss the implications.

(D. If it turned out that your group had to do a non-standard analysis to understand your data, such as controlling for an independent variable, defining separate sub-dimensions of the variable, or having to use the open-ended question or the general question as a criterion, insert a discussion of what you did here or prior to B and C, whichever seems most appropriate to what is going on.)

IV. **Results.**

A. **Univariate results for dependent variable**. 1) What do the frequency tables tell you about the distribution of opinion/behavior in your sample? What is the range of opinion? What is the typical opinion? Refer specifically to tables in the appendix, and give some examples or explain the basis of your answer. 2) For which closed-ended question(s) were people especially low on the variable? For which were people especially high? 3) What did you find interesting in the univariate results?

B. **Test of Hypothesis**. You will have either a correlation coefficient or a difference of means showing the relation between your independent variable and your dependent variable index (original or revised). If there is a single correlation, you may just present it in the text. If there is a difference of means, prepare a table to include in the paper (do not just cut out the table from the printout). Discuss this result, indicating whether your hypothesis is confirmed or rejected.

C. **Relation of dependent variable to other independent variables**. These will be correlations or differences of means. Briefly summarize what these show. (No more

than one sentence per independent variable; less is possible.) Instructions about presenting data as in B.

(D. Other analyses. Nothing else is required, but if you did something else for some reason, put the discussion here. For example, if you decided to control for a third variable.)

V. **Discussion** (NOTE: Write at least two paragraphs. Give at least some answer to every question. You do not have to give long answers for all four questions, but do give a serious and thoughtful answer to at least one. If your study interested you, discuss that the most. If your study seemed boring, discuss the "side" issues more.)

A. Explain the conclusions you draw from testing your hypothesis and from the frequency distributions. This is your chance to explain the wider significance of your work or to speculate a bit. In this section, you don't have to worry so much about sample bias, small sample, or what you did well, and instead talk about what was interesting to you and why. It should, however, be about the results of your survey, rather than some other topic.

B. Imagine this is a pretest for a larger study. Does it seem worthwhile to pursue this research? What changes should be made? Or, do you now think this is a dead end?

C. What did you learn from the process of doing this study? Does it change your opinion of survey research? Would you like to do this kind of research again?

D. Is there anything else that came up while doing the research that seems interesting or worthy of discussion or criticism?

VI. Appendices

1. Attach to the report a blank copy of your questionnaire which shows variable names and information about coding.
2. Attach the code sheet that was used for entering your data into the computer.
3. Attach the printouts of your frequency tables, correlation matrix, and reliability analysis. You do not have to recopy these tables – this will save busywork. But do write notes on the printouts to clarify variables or highlight items you are mentioning in the text. Also please remove extraneous paper to save weight when I have to carry this home.
4. Attach the original printouts for the tables you do have to recopy, as explained above, so I can be sure you understood what to copy.
5. Submit <u>all</u> the completed questionnaires when you submit the assignment. Make sure there are code numbers on each questionnaire to correspond to the identification numbers on the data sheet.

VII. Group process report. Pick the category that applies to you and answer the relevant questions. NOTE: YOU MAY NOT SHOW THIS REPORT TO ANY OTHER PERSON, AND YOU MUST SUBMIT THIS REPORT IN A MANNER THAT MAKES IT CLEAR OTHERS DID NOT HAVE ACCESS TO IT.

A. *No partner*. 1) How did you feel about working alone? Would you do it again, or would you prefer a group? 2) Tell me if there is anything I should know that might affect your grade or my ability to be fair in grading your work.

B. *Had partner, wrote separate papers*. 1) Compare you and your partner in the effort you put into the project. 2) Compare you and your partner in the extent to which you studied course materials and knew what to do for the assignment. 3) Who developed the questions? 4) Who prepared the tables from the printout? 5) Who figured out how to interpret the statistical results? 6) Did you start trying to work together before deciding to write separate papers? How far did you get? 7) Were there some things you found necessary to discuss in preparation for writing your papers? What? 8) How did the group process work out? Was it a positive or negative experience? Would you do things differently in the future? 9) Tell me anything else I should know that might affect your grade or your partner's, or that I should know to be fair in grading your work.

C. *Wrote joint paper*. 1) Do you stand by the paper as written, or is there something you feel should have been said differently? Any corrections you offer at this point will be factored into your grade. 2) Compare you and your partner in the effort you put into the project. 3) Compare you and your partner in the extent to which you studied course materials and knew what to do for the assignment. 4) Who developed the questions? 5) Who prepared the tables from the printout? 6) Who figured out how to interpret the statistical results? 7) How did you go about getting the writing done? 8) How did the group process work out? Was it a positive or negative experience? Would you do things differently in the future? 9) Tell me anything else I should know that might affect your grade or your partner's, or that I should know to be fair in grading your work.