ISSUES IN HUMAN SUBJECT TESTING IN CARTOGRAPHY AND GIS

Judy M. Olson, Professor Emerita
118 Geography Building
Michigan State University
East Lansing, MI 48824-1117
olsonj@msu.edu

Abstract

Introduction

Many studies in cartography and GIS use human subjects. In addition, we draw from human subject research in other fields, including psychology, human factors, and usability.

Objective

My objective is to present salient issues that will generate discussion and contribute to the dialogue on research involving human subjects.

Methodology

The set of issues was developed based on past experiences in reading, doing, evaluating, and guiding research. Issues vary in nature and are not intended to be exhaustive.

Results

Major observations include: 1) Journal review should include review of test instruments when feasible. 2) It would be helpful for test materials to be available to readers of published research. 3) Question design is crucial to the validity and meaning of the research. 4) The necessity of screening by an institutional review board can be both positive and negative. 5) Students and other young scholars need to be aware of the requirement for IRB approval for publication in some journals. 6) Multiple methods (quantitative and qualitative) are necessary for understanding human interaction with maps. 7) There are good arguments for balancing "good research" and "doing something".

Conclusion

There is no shortage of issues associated with human subjects research. These issues may help generate discussion that encourages and perhaps even improves human subject research in cartography.

Introduction

Many studies in cartography and GIS use human subjects. In addition, we draw from human subject research in other fields, including psychology, human factors, and usability. I have conducted studies with human subjects; read, reviewed, and edited such studies; supervised students conducting them; and served on an institutional review board that evaluated proposals.

Objectives

My objective is to present salient issues that will generate discussion on research involving human subjects. These issues are not likely to surprise researchers. On the other hand, most presentations on human subjects research projects focus on the specific substance of the project and not on the myriad of decisions and steps that go into the design and execution of it, decisions and steps that will be confronted by successive researchers.

Methodology

I have used my past experiences to develop the set of issues. Other researchers writing a paper of this title might have chosen issues that differ from these. But they will also likely resonate with at least some of those presented here. The issues vary in nature and are not intended to be exhaustive. The organizing of experiences and the thinking that has gone into this project creates a set of underlying issues, the discussion of which can potentially have a positive effect on this genre of research.

Results

The issues are presented here in no special order. Author biases will be apparent in what is said about them, and other researchers are at least as qualified to have opinions that may differ. Examples have all been constructed based on things in actual projects (cartographic and otherwise) but are distinctly fictionalized. Some of the negative examples sound naïve, but they all have some basis in the reality of legitimate projects that may in other ways be well thought out, well constructed, and worthy studies.

The need for journal review of test instruments

When a paper is reviewed for publication it is difficult to judge quality of conclusions without seeing test instruments. Impeccable statistical treatment of data is not sufficient for valid conclusions; the test itself also has to have been sound.

Without reviewing test instruments, serious flaws can be hidden in summarized results. It is not uncommon, for instance, to see that a question was asked "about" some

particular aspect of a map without the authors stating the actual questions used. Statistical results are presented in support of a conclusion "about" that aspect.

When test instrument materials, partial if not entire, are published as part of the article, the reviewers have had ready access to them, and there are examplary cases (Brewer and Pickle 2002; Coltekin and others, 2009). It is doubtful the instrument has been available to journal reviewers when not so published.

Availability of test materials for readers of published research

Readers of journal articles also have an interest in seeing test instruments and should, when possible, be able to review them in the same manner as the journal's reviewers. Further, the researcher who builds on previous studies benefits from being able to read, and perhaps even experience, the test instruments used. What should be our standards of availability? It seems reasonable that journals should require, where feasible, that test instruments be published as part of the research paper. Some materials simply cannot be included in the printed or even the online journal for reasons of length or the technology that was used for delivery. A test that has had subjects sitting at the controls and "driving" in a virtual space using a patented simulator, for example, could not be published in full in a printed journal. Test materials from projects such as this might be made available at least in very limited form on the Internet, assuming permission can be gained for needed components when appropriate. To the extent feasible, readers should be able to count on materials being available, in as full a form as feasible and reasonable, either online or by request, for a reasonable length of time.

It would be self-defeating to have such strict requirements for availability that publication of good research would be discouraged. On the other hand, some researchers have set good examples in putting their research instruments online when they cannot be incorporated into the printed journal article itself (e.g., GeoVista 2000-2002, Mitbø 2006). We would do well to encourage wide emulation.

Question design

Question design is so central to good human subjects testing that one might think good questions are universal in such research. However, we have probably all found ourselves trying to be cooperative in responding to research questionnaires but finding items on them that are structured illogically and are needlessly difficult or impossible to answer. We may even have refused to participate in what should have been worthy research but was not because of the ill-conceived questions they contained.

Consider a questionnaire with a scale of this pattern: 1—never, 2—infrequently, 3—neither yes or no, 4—often; 6—always. The scale labeling is flawed, as "neither yes or no" does not fit a frequency scale. Even worse, if questions do not begin "How

frequently…" but rather with "How do you…" or "The X is sometimes Y," they cannot be answered with an indication of frequency.

Another flawed type of question is to ask how confident the subject is that the answers given on a test were correct. What does the person answer if s/he was highly uncomfortable on many of the questions but not all? Should an average (in the general sense of an arithmetic mean) be estimated? Or does it really mean "How confident were you on most questions" (something more akin to the mode)? Perhaps the subject thinks s/he thinks about 70% right; is that 70% confident?

I have also seen questions that elicit direct answers to research questions instead of eliciting responses from which the researcher can draws appropriate conclusions in answer to the research question. Consider this hypothetical question: "Is the symbol on this map that represents quarries a good one?" We may find that 92% of respondents answer "yes" and with a reasonable number of subjects, that is certain to be statistically significant. But what does it mean? More appropriate tasks might have been: have subjects search for the symbol on a realistic map, ask what the symbol meant without showing a legend, test whether subjects remembered what the symbol stood for after seeing it in a legend, ask which of several pictures is likely showing the feature represented by the symbol, etc. Then *the researcher* could draw the conclusion about whether it is a good symbol. We have no idea what criteria the subjects used when answering yes or no to the direct question. We cannot even be sure they really thought it was a good symbol, as a meaningless yes-no question may well have been biased by any number of extraneous influences—trying to encourage or impress the reseacher, finding the symbol "cool" for some reason, feeling good about some event that just happened and transferring the good feeling to answering the question, familiarity with the symbol, or perhaps such unfamiliarity with the feature it stands for that one is guessing it is probably a good symbol for those who do know more.

Many cartographic research studies have impeccably good questions that would serve well as models of thoughtful question design. The projects referred to in the references are all in that category. If the exemplary research questions were available to readers of research papers as a matter of policy as suggested in the paragraphs above, less-experienced researchers would have positive examples of questions from which to learn.

Institutional review boards

Most, if not all, countries of the world have laws or policies regulating the use of human subjects, and many require prior review of instruments and procedures. A common term for the boards that do those reviews (in universities and other institutions that conduct research) is "institutional review board" (IRB). Does the requirement to receive approval from that board affect what is done in cartography and GIS? Review boards do not intentionally discourage research; but do students, faculty, or other researchers sometimes avoid research with human subjects because they perceive the review as a

hurdle? Are we neglecting pilot testing because changes at the end of the pilot test likely require further review? And although review boards protect subjects, not guarantee high quality research, is it ethical to allow human subjects to be recruited for bad research even if it won't physically or mentally harm them? Is the requirement for prior review increasing research quality because researchers know that someone will be scrutinizing the project or because having to submit in advance gives a few days for reflection, when potentially ruinous typos or inconsistencies can come to light?

The answers to these questions likely varies widely from one researcher to another. Some researchers resent such review with a passion, and some are either philosophically highly in favor or have perhaps recognized distinct benefits to their own research. But most researchers are probably just resigned to the process and see it perhaps in the same light as having to taking grant proposals through institutional approval—drudgery, but it is part of the research process.

At the very least, students headed for research careers that will include human subjects testing need to learn respect for the process and how to get through it as smoothly as possible. It is not always required that students obtain IRB approval for projects done strictly as a requirement within a course (no publication of results). But allowing students to do that simply insulates them from the process at a time when they should have ready sources of mentoring on the process, and the avoidance gives them a limited view of what is involved in really doing human subjects research.

IRB approval requirements for publication

Many journals have no policy with regard to whether a study has been reviewed and approved by the committee charged with protecting human subjects. For studies conducted in cartography and GIS it is probably unusual for subjects to have bee in any danger, and we likely trust that the researcher has complied with institutional requirements for IRB approval. When researchers publish in journals in other fields, however, there may well be a requirement for approval. It is unlikely that students writing theses and dissertations will be shut out of those publications because universities are making it increasingly clear that theses and dissertations can be rejected (and degrees delayed or denied) if human subjects have been used without approval. If a shortcut has been taken, whether by students in a class or faculty who do not understand the rules, some outlets for publication will simply not consider the article.

The need for multiple methods (quantitative and qualitative)

Our field has come a long way in accepting a wide variety of research methods, and some studies are exemplary in the use of multiple methods. Having reviewed a number of student projects outside our field that stick tightly, but not necessarily appropriately, to a single method, the virtues of combining multiple methods probably needs reinforcement in our own field as well.

Ways in which to incorporate multiple methods includes using open-ended questions to explore with subjects before launching a more tightly-controlled statistical experiment, using interviews and focus groups along with controlled experiments, and simply asking for reflection from some or all subjects following an experiment. Again, many researchers do use those methods, but the methods are also ignored in many studies.

Balancing "good research" and "doing something"

The ability to develop research projects of any kind comes with practice and time. Sending students out to "do something" with human subjects is a tempting, and to some degree valid, educational exercise, but one may be burdening subjects with useless activity. One of the practices that helps is "peer piloting." As part of a class, students are required to take one another's tests and critique them. Peer evaluation of the critiquing itself can be useful with questions such as "Did X give you helpful comments on your research" and "Was X receptive to your comments about his/her research". The point can be made with the students that both doing a good job in critiquing another's project and making good use of critiques received on one's own project are positive reflections on them.

Such required peer piloting and critiquing should be especially helpful at the upper undergraduate and beginning graduate levels and helps bring "doing something" closer to doing "good research." It also means that pilot runs can be done before seeking IRB approval because that part of the research is strictly within the class and for the benefit of the students within it.

Conclusion

There is no shortage of issues associated with human subjects research. These points are intended to engage cartography/GIS scholars in discussion.

References

Brewer, C.A., and L.W. Pickle, **2002**, "Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series," *Annals of the Association of American Geographers*, 92(4): 662-681.

Coltekin, Arzu, Benedikt Heil, Simone Garlandini, and Sara Irina Fabrikant. 2009. Evaluating the effectiveness of interactive map interface designs: a case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science* 36:1 (Jan): 5(13).

GeoVista. 2000-2002. Digital Government Quality Graphics, #6. Cluster Identification Experiments, <http://www.geovista.psu.edu/grants/dg-qg/pubs.html>, accessed July 2009.

Mitbø, Terje. 2006. Personal correspondence referencing a test instrument from earlier work with Keith Clarke and Sara Fabrikant <http://csiss.ncgia.ucsb.edu/resources/time/>, accessed July 2009.