Developing a Design Research Program Protocol

Help from the IRB (Institutional Review Board)

Mike Zender

Professor, Director of Graduate Studies, University of Cincinnati School of Design, mike.zender@uc.edu

Abstract: Nascent design research has not seen the development of many research programs that pursue a consistent line of inquiry over multiple studies. This paper proposes the qualities that define a research program in design, and describes how the individual methods that had been used in my symbol/icon design research: subject drawing; evaluation interview; comprehension estimation; comprehension survey; were integrated in an emerging icon design research program. An Institutional Review Board (IRB) protocol was the instrument that defined the research program. In addition to the obvious benefits of creating one comprehensive IRB protocol covering many studies over creating a variety of IRB protocols for individual studies, the paper discusses how creating an IRB-approved design research program protocol refined the design research methods used, defined areas of consistency (methods, subjects, data analysis) and variability (content, subjects, contexts), and insured that these would applied systematically to a similar palette of evolving questions over time, thus building knowledge. While the aim of the icon research program is better knowledge of icon design, the broader aim of this paper is to improve the rigor of design research as a valid discipline by advocating more design research programs.
Key words: research program, research methods, research protocol

1. Design Research Programs Needed

Numerous authors have called for design to mature from its crafts heritage. Sharon Helmer Poggenpohl cited Krippendorf, Owen and Buchanan's calls for change in her call for the building of design into a discipline that includes "a regimen that develops or improves a skill." (Poggenpohl, 2009, p. 8) A skill-developing regimen suggests some consistency of action and thought that can drive a knowledge-building process. Much design research to date has been one-off studies that by their nature limit opportunities for knowledge growth, stunt the refinement of methods, and fail to provide the substance and time for community verification by other researchers and scholars. An exclusive diet of isolated research studies may be retarding development of design as a discipline. An obvious way to overcome the limitations of single study research is to develop research programs that focus on a consistent topic for an extended period of time.

I'm not the first to call for research programs in design. While research programs in the sciences often cover multiple researchers across several institutions focused on a single topic that's grounded in established theory, the research program I'm describing in this paper is more modest:

a consistent topic,
an established palette of related methods,
longevity,
aimed building a knowledge base for the topic,
for verification by others.

While it would be wonderful for design to explore interesting questions surrounding proven theories as in the sciences, at this point there is not much design theory that has been validated by the wider design community to be true, that is, applicable in a wide variety of circumstances with predictable results. This makes it difficult to do a multi-site research program in design at this time. There are of course exceptions. Josef Albers' exemplary research program in color, concluded 50 years ago this year, produced knowledge that has been proven true over and over again in design foundation courses all over the world. (Albers, 1963) But as often, the exception proves the rule. Fifty years is a long pause between research programs if a discipline wants to build knowledge for advancement!

The research program I describe here has several features that might apply to any research program. It's unified by a single purpose – the design of icons; built on a common background of findings – perception and cognition; supported by resources - funding, facilities, trained team members; uses an integrated set of methods – creation and comprehension methods; adheres to applicable ethical standards – Institutional Review Board (IRB) protocol review process; and it handles and analyzes data rigorously.

It so happens that this list of program features parallels the structure of my institution's (University of Cincinnati – UC) IRB protocol template. In the United States IRB's were established in the 1970's in response to lax research standards and occasional unethical human subject research practices outlined in the Belmont Report. (full title: *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, Report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*.) The aims of University of Cincinnati's Office of Research Integrity, in which UC's IRB operates, are to insure ethical and quality compliance while facilitating research. The IRB itself is the board that reviews research proposals and oversees compliance with published guidelines particularly for human subjects.

At UC the IRB reviews all human subjects research protocols. A research protocol can define a research study, or research program. A research protocol both grows from a research activity, often from pilot studies, and governs the conduct of future research activities. A research protocol is a design object in that it governs the production of activities and products: the research activity and its outcomes. It's ironic that so few designers design research studies, and that so many designers seem allergic to writing research protocols, since creating a research protocol is an exemplary design activity. The University of Cincinnati requires an IRB approved protocol for any human subject research that aims to produce generalizable knowledge. Human subject research includes everything that deals with human subjects from the most invasive surgeries to the most casual interviews. Generalizable knowledge is any finding intended to cover numerous cases. This research definition is confusing in design, and the source of some debate within design research, because it means that activity to gain knowledge to improve a single product or service, often called research in design, is not considered research at UC (or most other places). Research "to design" is not research at all because it does not apply broadly but only to a single

case. Such "research" is exempt from IRB oversight using the "product improvement" rational. Since I've argued above that design lacks generalizable knowledge, and since I'm seeking to produce generalizable knowledge for design that can serve to develop principles to guide design work across projects and disciplines, I have placed my work within the IRB definition of research and submitted research protocols to the IRB for approval. While writing research protocols to IRB places my research activity in the context of drug trials and death penalty studies, I've found that being party to research in other disciplines whose research paradigms are more mature than design's has improved the structure and attention to methodology of the design studies described below.

2. Icon Research Program Example

I noted above some proposed qualities of a research program as opposed to a single study:

a consistent topic,
an established palette of related methods,
longevity,
aimed building a knowledge base for the topic,
for verification by others.

To these mechanical attributes I would add two important qualities of a successful research program:

significant topic area,
narrow focus.

I propose these because it is wasteful to squander research resources on an insignificant topic. While this may seem obvious, there are pressures in academia to publish nearly anything that might lead to advancement, a degree, or tenure. All you have to do is read several journal article proposals or graduate theses to realize how trivial and useless some topics are. A significant topic would apply broadly and have a deep impact on, or outcome in, practice or theory formation.

The flip side of the significant topic is a narrow focus. This is necessary to limit scope so that something can actually be studied in a human lifespan. Again, all you have to do is review graduate students' research proposals to see how many are so ambitious that it would take many lifetimes to answer. I'm reminded of the cliché answer of the beauty contestant in answer to the question: "What do you want to accomplish?" answer: "World peace!" That's a significant but very broad topic.

My approach to the "world peace" quandary is that I want to understand how visual communication works through understanding how icons work; what makes some icons successful and others failures. This is narrow enough to study, but broad enough to be significant when one considers that principles for designing icons might be applied to designing successful symbols more broadly, and that nearly all communication design uses symbols.

My icon design research began in 2004 with a funded research study by Procter & Gamble to design icons that could communicate product attributes across language barriers. This study's findings were reported in *Visible Language* 40.2. (Zender, 2006) At about the same time I began a study with colleague Dr. Keith Crutcher to use an icon system to communicate concepts related to Alzheimer's disease extracted from published papers in a non-verbal icon concept map display. This work was funded by the University Research Council and published in *Visible Language* 41.1. (Zender & Crutcher, 2007) In this study we used both a think-aloud protocol where domain experts reviewed prototype icon displays, and an on-line survey to assess effectiveness. What followed was an increasing number of overlapping symbol/icon studies, using similar research methods, to design icons and icon systems. One study developed and evaluated an icon system for communicating concepts in epilepsy using an think-aloud comprehension test and interview method; another study was for an icon system to communicate

across language barriers for healthcare facilities sponsored by the Society for Environmental Graphic Design (SEGD) and Hablamos Juntos that used a comprehension estimation survey method (ISO, 2007); yet another study applied the Hablamos Juntos healthcare icons in a rural health clinic in Tanzania that required two trips, one using comprehension estimation and comprehension surveys, and the subsequent combining drawing something, comprehension estimation, and comprehension survey methods; still another evaluated an icon-based visual story to communicate scripture stories in Kenya using a novel narrative evaluation and user think-aloud comprehension interview method; another designed an icon system to help medical students memorize pharmaceutical interactions; and various other current studies create and evaluate icons and symbols for various contexts primarily in medical communication: from COPD to HPV.

All of these studies focused on icon design. Most used a separate IRB protocol for each study. As study parameters or personnel would change I would write a protocol modification to cover any change in the specific question, or subject profile, or icon content. For example, when we took the 54 Hablamos Juntos icons to Tanzania we found that most of the icons were not understood at acceptable levels (85% correct comprehension) and theorized that one reason might be that certain medical concepts were simply not common knowledge among Tanzanians. So we designed a follow-on study in the United States to measure differences in comprehension between those with normal and advanced medical literacy and compared these results to results from Tanzania. This required a new IRB protocol. In another case, we identified four troublesome Hablamos Juntos icons and hypothesized that removing a particular symbol from the icon would degrade its comprehension significantly. Because of how the original protocol was written, this required yet another IRB protocol modification. The protocol design and modification process I was using became inefficient for everyone: research team and IRB staff alike.

3. Protocol for Research Program

The answer to this waste and frustration was to design a research protocol defining a research program that would include a number of individual research studies. The research program was to study icon design in a variety of contexts and applications, addressing numerous questions and sub-questions. To accomplish this I wrote, with the help of Claudia Norman a UC research advisor, a research protocol that integrated consistency and variability.

3.1 Consistency

The consistent features of the research program protocol were:

single broad purpose,
consistent palette of methods,
consistent (broad) population definition (inclusion/exclusion criteria),
consistent palette of recruitment methods,
consistent consent procedures,
consistent data handling and analysis.

The title of the study “Symbol/Icon Design and Comprehension” had a single broad purpose:

> “… to design and measure comprehension of symbols/icons that may be used in a variety of settings.”

Over several years of study on previous icon research projects, various research methods had been used, adapted (in some cases from ISO and ANSI standards), or invented. To achieve the purpose of this research program four separate research methods were integrated and described:

1. Subject Drawing

2. Comprehension Interview
3. Comprehension Estimation Survey
4. Comprehension Survey

Some methods were conducted on paper, some on-line, and some in both media

The first method is subject drawing:

1. Subject Drawing

Subjects are first given an instruction sheet describing how to properly answer the questions. Then they begin the instrument. An instrument page has one or more referents (the icon/symbol's intended meaning) on the page. The subject is then asked to draw the most representative picture that comes to mind when they think of that referent. When the subject has drawn something for each referent, the instrument administrator thanks the subject for their participation and the subject is free to go.

The Subject Drawing test is one we developed based on Steven Kosslyn's theory of mental imagery. (Kosslyn, Thompson, & Ganis, 2006) Kosslyn describes what we have called "brain icons" – simple visual representations in visual memory of familiar objects. Everyone has and uses these brain icons in the process of seeing and identifying familiar objects. We explored this method in Tanzania and have tested it in various countries to successfully identify common attributes of people's mental images. My family uses software to play the game "DrawSomething" which uses the same idea. In the research protocol, the Subject Drawing method was prescribed for used in early stages of the symbol/icon design process. Subjects' drawings are collected and visually analyzed for similarities and patterns of objects and point-of-view. The most common objects and viewpoints represent the brain icon people hold in their memory. The designer uses this data to guide his or her symbol/icon design.

The next method is Comprehension Interview:

2. Comprehension Interview

Subjects are first given an instruction sheet describing how to properly answer the questions and asked permission to record the interview. Then they begin the interview. An interview page has one or more symbols/icons with each symbol/icon. The subject is shown the interview page and each subject is asked a series of pre-defined open-ended questions. The kinds of questions being asked on these instruments will be:

What things do you see in this picture/symbol/icon?
How would you interpret this picture/symbol/icon?
Is there anything in the picture/symbol/icon you find confusing?
What would you change to make this picture/symbol/icon better/clearer?
How many people do you think will understand this picture/symbol/icon?
Exactly what do you think this picture/symbol/icon means?
What action would you take in response to this picture/symbol/icon?

The subject expresses an answer for each question in their own words. The interview administrator will record the comments and make necessary notes. When the subject has completed the interview they hand the interview instrument to the administrator and are free to go. In-person activities will take up to an hour and a half, depending on the length of the instrument.

The answers are scored, and the text of incorrect answers is analyzed to give feedback to the designer for improvement. The qualitative, open-ended nature of the Comprehension Interview method gives maximum variety and freedom of feedback. It's hard to score, but useful when you are not at all sure why something is not working/working and/or how to improve it.

This method is often followed by Comprehension Estimation, as defined in ISO and ANSI standards for safety symbols:

3. Comprehension Estimation Survey

> Subjects are first given an instruction sheet describing how to properly answer the questions. Then they begin the survey. A survey page has two or more candidate symbols/icons for a single referent (the icon/symbol's intended meaning) arranged around a central written expression of the referent. Subjects are asked to estimate "How many people will understand this symbol?" For each symbol, subjects either indicate in a percentage, or indicate a number representing "none – some – half – most – all". Each group of candidate symbols/icons has the same questions. When the subject has completed the survey instrument they hand it to the administrator or click "Finished" for an on-line survey, and are free to go.

We have found that estimating percentages is both too precise in an estimation context, and confusing for many people who have trouble calculating a 20% tip. However, the ISO standard 9186-1 calls for use of the percent approach so we included it. As an alternative, we have developed a scale "none – some – half – most – all" illustrated by 5 small human icons and ask subjects to circle how many people they think will understand each icon. This approach is more successful, less confusing. The answers are collected, tabulated and the results given to the symbol/icon designer to guide design of the most effective symbol/icon. The designer will compare the components/symbols/illustrations of the high scoring approaches for similarities, while avoiding components/symbols/illustrations/traits common to low-scoring symbol/icons.

The final method is Comprehension Survey, again, based on ISO and ANSI standards:

> 4. Comprehension Survey
> Subjects are presented with an instruction page describing how to properly answer the questions. Then they click to the next page and begin the survey. The survey page gives a brief scenario explanation of the context for the symbols/icons and has one or more symbols/icons with each symbol/icon having the same set of questions. The kinds of questions being asked on all these instruments will be:
> *Exactly what do you think this picture/symbol/icon means?*
> *What action would you take in response to this picture/symbol/icon?*
> The subject will indicate their answer for each question by writing/typing their answer in their own words. When the subject has completed the survey instrument they hand it to the administrator, or click "Finished", see a "Thank You" screen for the on-line survey and are free to go.

Survey answers are scored by a minimum of 3–5 scorers. Each referee works from a score sheet that defines acceptable correct answers and assigns a score of correct; partially correct; incorrect; or fatal. The text of incorrect answers is analyzed to give feedback to the designer for improvement. Designers analyze the open-ended answers to low-scoring symbols to gain particular insight for correcting/improving them. We wrote the protocol so that these four methods may be used sequentially, concurrently, or separately to address different contexts or questions. For methods 1 and 2 we described using about 8–10 subjects for each study. For method 4 we described needing a much larger sample: 90–100 subjects at a minimum.

For the research program protocol we defined very broad subject inclusion/exclusion criteria to allow the maximum flexibility:

> Volunteer participants will be male and female adults, 18 years of age or older. Only minors will be excluded. Included subjects may be undergraduate students. However, any students enrolled as study subjects will not be students of a research team member (the PI or the study administrators).

The last point about excluding students of the research team was important because the study is done in an academic context where issues of power and coercion are in play between faculty and students and even upper classmen and lower classmen. Coercion was a significant concern of the Belmont report and a major issue with the IRB.

The recruitment methods for these subjects were standard and common across all methods and study questions: advertisement, public announcement, poster, flyer, email solicitation. One part of the recruitment description process said:

Depending on the project context, research administrators will personally invite people by oral presentation, and/or through word-of-mouth, or by posting a flyer, or by placing an advertisement in an appropriate media or on-line venue, or by email to an appropriate list (with permission of the list owner who also has documented permission to send email to list members). Various projects targeted toward different populations will use different combinations of recruitment methods. Any recruitment method may be used for each project. Instrument administrators may maintain a sign-up sheet for responses from potential subjects. This sheet may list phone numbers or email addresses given by the potential subjects for contact purposes. The sheet will not be posted publicly. Those excluded (minors or students of study administrators) will be thanked for their willingness to participate on this important study and informed that only adults and non-students of study administrators are eligible for this study.

Our IRB prefers that recruitment be general and broad with the volunteer subject taking the initiative to come forward to volunteer rather than individual pressure being brought to bear on a single individual. That said, it's not uncommon for student researchers to invite their friends and even family members to participate in a study. This practice is not unacceptable, but it is not preferred.

Data handling and analysis for the research program protocol were also fairly standard, but with a specific description of data handling and analysis for each method:

When the instruments are completed, results will be scored using a standard scoring sheet indicating correct answers. One or more independent scorers will score each instrument response using the score sheet.

Subject Drawing methods are descriptive and will be analyzed visually for common features. Common features of subjects' drawings are summarized, analyzed for their fit to referent meaning, and synthesized into a preferred approach approximating the image most commonly held in most participants' minds. This synthesized image approach is then used to guide symbol/icon design for a Comprehension Estimation Survey.

Comprehension Estimation Surveys symbol/icon scores: none; some; half; most; all; will be totaled and described using descriptive statistics (a mean for example) and inferential statistics (T test for example). These results will be analyzed and the most successful symbol/icon approaches identified. Common features of both the most and less successful symbols/icons will be identified and analyzed. Unsuccessful symbol/icon features will be avoided, and more successful features will be noted and synthesized into a final symbol/icon for further testing (Comprehension Interview or Comprehension Survey), or for final development.

Comprehension Interview and **Comprehension Surveys** will be scored by a minimum of three scorers. Each symbol/icon will have one or more possible scores assigned depending on the instrument. Typical scores are: correct; partially correct; incorrect; fatally incorrect. An expert panel will adjudicate scoring questions and disputes. Scores given to each of the symbols/icons will be entered into a spreadsheet and then re-checked twice. Scores will be totaled and described using descriptive statistics (a mean for example) and inferential statistics appropriate to the project. This data will then be analyzed to determine which symbols/icons scored above 85% and which did not. Symbols/icons with a score above 85% will be considered successful, symbols/icons scoring less than 85% are considered to have problems with comprehension and thus needing redesign. In addition to scoring, open-ended Comprehension Interview and Comprehension Surveys responses for each symbol/icon will have the language of their answers analyzed using natural language processing to identify patterns. These patterns provide clues to the cause of the success and failure of the symbol/icon. The language analysis and the score results are described using visualization to gain further insight. The verbal clues, the score results, and the visualization of these provide information on which to base strategies for design improvements. Designers will then apply this data to the redesign of unsuccessful symbols. These symbols/icons can be retested and redesigned until the 85% threshold of comprehension is achieved.

Throughout these processes we hope to identify principles for more effective symbol/icon design.

3.2 Variability

As you can see from the above, the consistent elements of the research program protocol allowed for considerable variability. The purpose of the protocol allowed for great variability of multiple projects to address a diverse and evolving set of research questions surrounding the single topic of symbol/icon design:

> The purpose of this research study is to design and measure comprehension of symbols/icons that may be used in a variety of settings.

By not specifying any content for the symbols/icons (such as pharmaceutical side effects), or even a content domain (such as medicine), we allowed the possibility for the research program to change specific content within general subject area of symbols/icons. This is very broad indeed. We also allowed for changing demographics of recruited subjects by keeping the inclusion/exclusion criteria broad. This will allow us to change the research contexts to anything of interest. We acknowledged with the broad data handling and analysis descriptions that while we generally would be reviewing the same kinds of data: about icons, that this data could have different specific features: those with average and advanced medical literacy for example, or those with and without English proficiency. Due to the nature of the different methods, we allowed for variable numbers of subjects (within limits):

> Number of participants
> Minimum and maximum number of participants
> Within this study a variety of research projects will examine symbols/icons in different contexts. For each project up to 500 subjects may be required. The total number of subjects in the research study will depend on the number of projects being done.
> Rationale
> Various projects will examine different stages in the symbol design process. Early design stages are focused on discovery of successful symbol options and require few subjects: 6 – 10 for example. Later design stages for evaluation and comparison use inferential statistics and require more subjects: 100 – 500 for example. Each project will use the number of subjects for sufficient balance between statistical reliability and ease of testing, in line with the anticipated effect size.

The IRB is concerned with use of resources including the exhaustion of the pool of subjects for studies. This is one area on which the IRB asked for clarification, wanting to know the upper limit of subjects for the entire study program. I gave the answer of 5,000, which was accepted. Finally, over time personnel will change. There is no mechanism to approve this in advance, so the research program protocol will still have to be modified as personnel are removed and added over time.

As noted previously, Claudia Norman, a faculty connected to UC's IRB office, helped in drafting the above IRB protocol for a the symbol/icon research program. Her insights into the concerns of the IRB were invaluable.

4. Conclusions

The research program protocol was approved and has been in use. Without doubt IRB interaction resulted in a better protocol by forcing clear descriptions of the methods to be used, describing the apt context for each method, and defining areas of consistency and variability. Overall the methods were a significant improvement over the previous protocols I had developed which seemed ad hoc, ND seat-of-the-pants research by comparison. In summary: working with my IRB has made my research better: more rigorous, more structured, more innovative since apt structure stimulates innovation.

Further, developing a research program protocol helped advance my research program! It helped me the define core purpose of my research program; it helped define subjects' parameters expressing what was consistent and what was variable in my subjects; it helped me to analyze methods and categorize them in relation to study types and to identify the advantages and deficiencies of each method; it enabled a series of related studies: related by consistency of subject with variety of features, related by consistency of method with variety of combinations; it insured consistent rigor across studies which in turn enables direct comparison of one study with another, facilitating accumulative learning.

As you can see, I'm a fan of the design discipline developing more research programs. I literally cheer the thought of various design research centers that focus on significant and strategic design questions and address them in a systematic way. Creating the protocol with IRB support helped me advance my line of research while reducing the burden of constant small modification requests. I hope that this paper helps stimulate other design research programs as well

5. Citations

Albers, Josef. (1963). *Interaction of Color* (revised and expanded paperback edition ed.). New Haven: Yale University Press.

Graphical symbols - Test methods Part 1: Methods for testing comprehensibility, 9186-1 C.F.R. (2007).

Kosslyn, Stephen M., Thompson, William L., & Ganis, Giorgio. (2006). *The Case for Mental Imagery*. New York, NY: Oxford.

Poggenpohl, Sharon. (2009). Time for Change: Building a Design Discipline. In S. P. a. K. Sato (Ed.), *Design Integrations Research and Collaboration* (pp. 3 - 22). Bristol, UK: Intellect.

Zender, Mike. (2006). Icon Systems for Global Non Verbal Communication. *Visible Language, 40*(2), 177-206.

Zender, Mike, & Crutcher, Keith A. (2007). Visual Language for the Expression of Scientific Concepts. *Visible Language, 41*(1), 23-48.