Coltekin, A. (2015). Mix well before use: Understanding the key ingredients of user studies . Workshop (position) paper -- Pre-conference workshop "Envisioning the Future of Cartographic Researhc" at the International Cartograhic Conference, ICC 2015, Curitiba, Brazil.

Mix well before use: Understanding the key ingredients of user studies

Arzu Çöltekin
Department of Geography,
University of Zurich
Zurich, Switzerland
arzu@geo.uzh.ch

ABSTRACT

There are many different approaches to conducting user studies, e.g., direct observations, interviews, focus groups, surveys, questionnaires, online experiments, and controlled experiments. None are perfect, but all are good. Each approach has strengths and weaknesses; therefore, a well-considered *combination* may offer the most reliable results by exploiting the strengths of one method to the benefit of the other(s). In this vein, we discuss a few selected (prominent) dichotomies relevant in experimental research design and give examples from cartography and geographic visualization. The broader discussion is relevant to all user research methods; however, we specifically reflect on what complements *controlled experiments* and offer insights from our experiments with *eye movement analysis*.

Author Keywords

User studies; cartography; geovisualization

ACM Classification Keywords

Experimentation; human factors

INTRODUCTION

Whether the study tackles an issue in e.g., medicine, astronomy or cartography, there are three basic ingredients in user studies: the *stimuli* (typically a “product”), the *tasks* (context), and the *participants* (human). Most experiments revolve around these three dimensions: we test a *thing* for a certain *purpose* and a certain group of *people* (e.g., a map for route planning for first time visitors of a city; a pill for chickenpox in children). Our results are a combined expression of the ingredients of these three ‘pillars’ and even the really tiny; seemingly insignificant details in each of these pillars can change the outcome of an experiment. Thus, our concerns for validity (and eventually generalizability) of our findings are directly linked to the limitations we introduce as we design the stimuli, the tasks, and choose our participants.

One way to assess the user research methods is to conceptualize them in terms of their *usability*. Taking the ISO definition of usability [1], we reflect whether our methods are effective, efficient and satisfying to the scientist who is trying to answer a specific question. An *online study* may be efficient (i.e., fast: we can get hundreds of responses in a week) and perhaps they would operate as they normally would (i.e., “ecologically valid”: instead of coming to an unfamiliar lab, perhaps feeling observed); but is it really good enough in terms of effectiveness (i.e., success, accuracy)? Since we have no control of what the user is doing, they could talk to someone else, search for the answer online, or take a call during the study. If this is the case, the success (or failure) will not be theirs alone. Furthermore, each participant will be using a different computer with a different processing speed, bandwidth, and screen. All these might affect the findings in profound ways in visualization studies. Similarly, we worry about *focus group* studies. What if people just want to agree with the moderator or with the first (or most dominant) person in the group (known as “acquiescence bias” [2], which applies to many other studies and this is why we need to take utmost care in how we word the questions) or choose what they already believe to be the good thing even if they are ill-informed (“confirmation bias” [3]). *Interviews* may suffer from deriving only ‘self-declared’ success with the visualizations [4], while *direct observations* clearly would depend very much on the observer [5]. If we, then, chose to run a carefully designed *controlled lab experiment*, is it perhaps *too* controlled? To be able to identify a cause-effect relationship, controlled lab experiments attempt minimizing confounding variables through making sure that only a single variable is modified at a time [6]. However, in most cases the necessary simplification of tasks and stimuli makes it very difficult to generalize (this problem is known as “experimental control vs. ecological validity” [7]).

These concerns have been voiced for all user study methods. Below we will present a few selected dichotomies on these concerns and discuss why we should consider using mixed methods to address them.

SOME DICHOTOMIES IN USER STUDIES

In the scope of this discussion paper, we selected a few dichotomies that we consider important for user studies. Note that this is not a comprehensive survey of all competing paradigms.

Top-down (hypothesis- driven or confirmatory) vs. bottom-up (exploratory) analyses

Most fundamental experimental science is (justifiably) conceptualized top-down, i.e., hypothesis-driven; meaning that we have an informed position as to why we ask the questions we ask. However, there are benefits in exploring the collected data bottom-up, as this may reveal unexpected patterns and lead to new hypotheses [8]. For example, in an eye movement study, areas of interest (AOIs) can be designed “top-down” for the design elements we intended to test. However, exploring the gaze patterns beyond the top-down AOIs to inspect where else participants were looking (“bottom-up AOIs”) may give interesting information on distractors or dominant design elements. In cartographic research where the visual stimuli cannot be entirely reduced to a single-variable item (as opposed to perceptual psychology research), this is especially valuable. It is important to be aware that top-down approaches are prone to type I errors (“false positives”), where the bottom-up approaches have a bigger risk for type II errors (“false negatives”), and clearly state which of these approaches were used in what part of the study [9].

Measuring attitude vs. behavior

Seemingly, what we say (attitude) often does not predict what we do (behavior). The concepts *naïve realism* [10] and *naïve cartography* [11] are expressions of this. At first sight, behavior might appear more interesting than attitude, but again, a mix of the two offers valuable insights. For example, in a comparative study, if the user performance does not differ, but they prefer one of them, we can study what makes one more desirable. Similarly if there is a mismatch between the attitude and behavior, this will inform our visualization (and interaction) design. Furthermore, adding extra behavioral measurements such as eye movement analysis might allow identifying strategies that participants themselves may not be aware of [11, 12].

Using qualitative vs. quantitative methods

Using qualitative or quantitative methods can create a division among some scientists [13, 14]. Despite the passion for one or the other, mixed-methods studies that utilize qualitative *and* quantitative methods are fairly common in user experiments, and this is for good reason. Qualitative methods can help with ‘why’ questions (as well as ‘how to’), while quantitative methods establish ‘how much’ (as well as ‘how many’ and ‘how long’), increasing our understanding of the phenomena we study [16].

CONCLUSION

For cartography and geographic visualization, as well as the wider scientific visualization research, controlled studies should be the central method, and be supported by additional research methods case-by-case. Since the controlled studies are time-consuming and expensive, we need to think about what can be recommended to practitioners (e.g., a ‘quick and dirty usability’ scale [17] or a ‘single score’ [18], or a short spatial ability test to predict map reading performance [19]?). However, for scientific research, where the goal is to understand the underlying reasons and identify “universal” patterns, controlled studies that are supported by additional methods (qualitative interviews, direct observations, eye tracking, etc.) are, in our understanding, most *useful*.

However, for our results to be ‘effectively’ useful, we need to become aware of the importance of the *effect size* [20] and make a habit of reporting it. Another important, seemingly lesser studied issue to be aware of is individual differences, e.g., in *spatial abilities* [20, 21], perhaps expressing themselves in map reading and other geographic tasks (see Figure 1) or *anxiety* [23]. If some individuals respond differently to our research methods, a mixed-methods approach may be beneficial also in this regard. Other psychological concepts such as *priming* [24] and *cognitive biases* [25] would similarly be best addressed through using more than one (carefully selected) method.

Perhaps a metaphor to “multiple-linked views” in visualization can be drawn; a “multiple-linked methods” approach allows approaching the studied phenomena from different perspectives; and may compensate for the lack of certain ‘usability’ aspects in one method by bringing in another that does not suffer form the same issue.

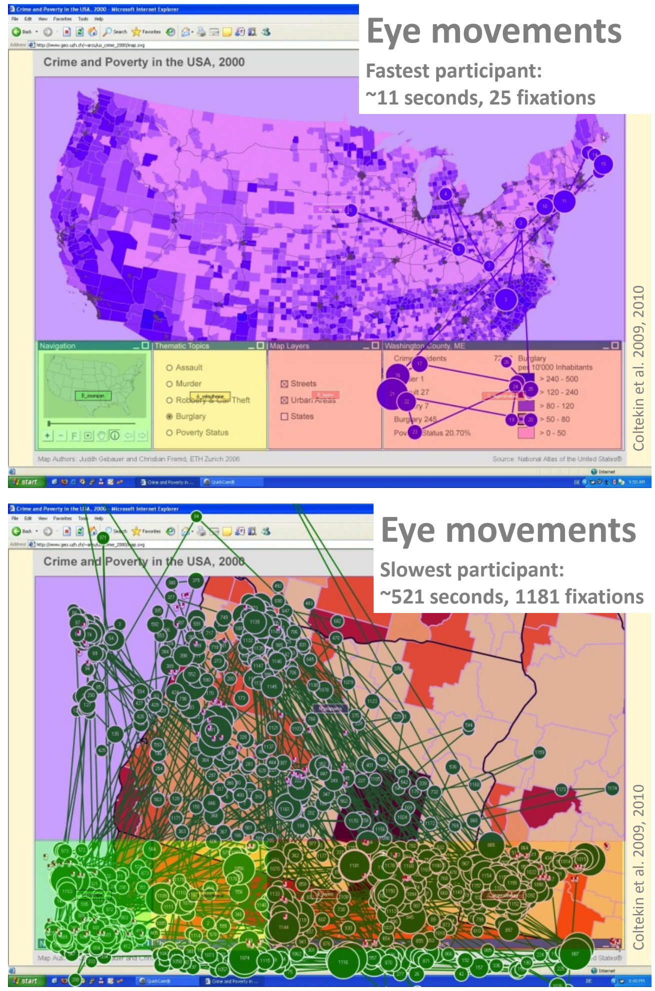

Figure 1. Individual differences between participants can be strikingly high.

REFERENCES

[1] ISO, "ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part II: Guidance on usability," *Int. Organ. Stand.*, 1998.

[2] P. J. Ferrando and U. Lorenzo-Seva, "Acquiescence as a source of bias and model and person misfit: a theoretical and empirical analysis.," *Br. J. Math. Stat. Psychol.*, vol. 63, no. Pt 2, pp. 427–448, 2010.

[3] C. R. Mynatt, M. E. Doherty, and R. D. Tweney, "Confirmation bias in a simulated research environment: An experimental study of scientific inference," *Quarterly Journal of Experimental Psychology*, vol. 29, no. 1. pp. 85–95, 1977.

[4] J. Potter and A. Hepburn, "Qualitative interviews in psychology: problems and possibilities," *Qual. Res. Psychol.*, vol. 2, no. 4, pp. 281–307, 2005.

[5] L. Baker, "Observation: A complex research method," *Library Trends*, vol. 55, no. 1. pp. 171–189, 2006.

[6] A. Blandford, A. Cox, and P. Cairns, "Controlled experiments," in *Research Methods for Human-Computer Interaction*, 2008, pp. 1–16.

[7] M. Cole, L. Hood, and R. P. McDermott, "Concepts of ecological validity: Their differing implications for comparative cognitive research," in *Mind, culture, and activity: Seminal papers from the Laboratory of Comparative Human Cognition.*, 1997, pp. xiii, 501.

[8] D. B. Kell and S. G. Oliver, "Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era," *BioEssays*, vol. 26, no. 1. pp. 99–105, 2004.

[9] R. G. Jaeger and T. R. Halliday, "On confirmatory versus exploratory research," *Herpetologica*, vol. 54, pp. S64–66, 1998.

[10] H. S. Smallman and M. B. Cook, "Naïve realism: Folk fallacies in the design and use of visual displays," *Top. Cogn. Sci.*, vol. 3, no. 3, pp. 579–608, 2011.

[11] M. Hegarty, H. S. Smallman, A. T. Stull, and M. S. Canham, "Naïve cartography: How intuitions about display configuration can hurt performance," *Cartogr. Int. J. Geogr. Inf. Geovisualization*, vol. 44, no. 3, pp. 171–186, Jan. 2009.

[12] A. Çöltekin, S. I. Fabrikant, and M. Lacayo, "Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. 10, pp. 1559–1575, Oct. 2010.

[13] A. Brychtova and A. Coltekin, "An empirical user study for measuring the influence of colour distance and font size in map reading using eye tracking," *Cartogr. J.*, p. 1743277414Y.000, Nov. 2014.

[14] O. Gelo, D. Braakmann, and G. Benetka, "Quantitative and qualitative research: Beyond the debate," *Integrative Psychological and Behavioral Science*, vol. 42, no. 3. pp. 266–290, 2008.

[15] K. R. Howe, "Getting over the quantitative-qualitative Debate," *American Journal of Education*, vol. 100, no. 2. p. 236, 1992.

[16] C. Rohrer, "When to use which user-experience research methods," *NN/g Nielsen Norman Group*, 2014. [Online]. Available: http://www.nngroup.com/articles/which-ux-research-methods/.

[17] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, 1996.

[18] J. Sauro and E. Kindlund, "A method to standardize usability metrics into a single score," *Proc. SIGCHI Conf. Hum. factors Comput. Syst. (CHI '05)*, pp. 401–409, 2005.

[19] A. K. Lobben, "Navigational Map Reading: Predicting Performance and Identifying Relative Influence of Map-Related Abilities," *Annals of the Association of American Geographers*, vol. 97, no. 1. pp. 64–85, 2007.

[20] R. E. Kirk, "The importance of effect magnitude," in *Handbook of Research Methods in Experimental Psychology*, 2008, pp. 83–105.

[21] M. Hegarty and D. A. Waller, "Individual differences in spatial abilities," in *The Cambridge Handbook of Visuospatial Thinking*, P. Shah and A. Miyake, Eds. Cambridge: Cambridge University Press, 2005, pp. 121–169.

[22] D. H. Uttal, N. G. Meadow, E. Tipton, L. L. Hand, A. R. Alden, C. Warren, and N. S. Newcombe, "The malleability of spatial skills: A meta-analysis of training studies.," *Psychological Bulletin*. 2012.

[23] M. W. Eysenck, N. Derakshan, R. Santos, and M. G. Calvo, "Anxiety and cognitive performance: Attentional control theory," *Emotion*, vol. 7, no. 2, pp. 336–353, 2007.

[24] D. L. Schacter and R. L. Buckner, "Priming and the brain," *Neuron*, vol. 20, no. 2. pp. 185–195, 1998.

[25] A. Wilke, "Cognitive Bias," in *Encyclopedia of Human Behavior*, 2012, pp. 531–535.