Full text available at: http://dx.doi.org/10.1561/1100000043

Some Whys and Hows of Experiments in Human–Computer Interaction

Full text available at: http://dx.doi.org/10.1561/1100000043

Some Whys and Hows of Experiments in Human–Computer Interaction

Kasper Hornbæk

University of Copenhagen
Denmark
kash@diku.dk

the essence of knowledge

Boston – Delft

Full text available at: http://dx.doi.org/10.1561/1100000043

Foundations and Trends® in Human–Computer Interaction

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is K. Hornbæk, Some Whys and Hows of Experiments in Human-Computer Interaction, Foundations and Trends® in Human–Computer Interaction, vol 5, no 4, pp 299–373, 2011

ISBN: 978-1-60198-682-5
© 2013 K. Hornbæk

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/1100000043

Foundations and Trends® in Human–Computer Interaction
Volume 5 Issue 4, 2011

Editorial Board

Editor-in-Chief:
Ben Bederson
Human–Computer Interaction Lab
University of Maryland
3171 A. V. Williams Bldg
20742, College Park, MD

Editors
Gregory Abowd (Georgia Institute of Technology)
Batya Friedman (University of Washington)
Jon Froehlich (University of Maryland)
Jonathan Grudin (Microsoft Research)
Jason Hong (Carnegie Mellon University)
Juan Pablo Hourcade (University of Iowa)
Karrie Karahalios (University of Illinois at Urbana-Champaign)
Gary Klein (The MITRE Corporation)
Joe Konstan (University of Minnesota)
Chris North (Virginia Tech)
Yvonne Rogers (University College London)
Orit Shaer (Wellesley College)
Desney Tan (Microsoft Research)
Kentaro Toyama (University of California, Berkeley)
Jacob Wobbrock (University of Washington)

Full text available at: http://dx.doi.org/10.1561/1100000043

Editorial Scope

Foundations and Trends® in Human–Computer Interaction will publish survey and tutorial articles in the following topics:

- History of the research Community
- Design and Evaluation
  - Ergonomics/Human Factors
  - Cognitive engineering and performance models
  - Predictive models of interaction
  - User-centered design processes
  - Participatory design
  - Graphic design
  - Discount evaluation techniques
  - Design and interaction
  - Ethnography
- Theory
  - Models of cognition
  - Empirical methods of evaluation
  - Qualitative methods of design and evaluation
- Technology
  - Programming the graphical user interface
  - Input technologies
  - Output technologies
- Computer supported cooperative work
  - History of CSCW in HCI
  - Organizational issues
  - Online communities
  - Games
  - Communication technologies
- Interdisciplinary influence
  - The role of the social sciences in HCI
  - MIS and HCI
  - Graphic design
  - Artificial intelligence and the user interface
  - Architecture and the role of the physical environment
- Advanced topics and tends
  - Information visualization
  - Web design
  - Assistive technologies
  - Multimodal interaction
  - Perception and the user interface
  - Specific user groups (children, elders, etc.)
  - Sensor-based or tangible interaction
  - Ubiquitous computing
  - Virtual reality
  - Augmented reality
  - Wearable computing
  - Design and fashion
  - Privacy and social implications

Information for Librarians
Foundations and Trends® in Human–Computer Interaction, 2011, Volume 5, 4 issues. ISSN paper version 1551-3955. ISSN online version 1551-3963. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/1100000043

Foundations and Trends® in
Human–Computer Interaction
Vol. 5, No. 4 (2011) 299–373
© 2013 K. Hornbæk
DOI: 10.1561/1100000043

now
the essence of knowledge

Some Whys and Hows of Experiments in Human–Computer Interaction

Kasper Hornbæk

Department of Computer Science, University of Copenhagen, DK-2300, Copenhagen S, Denmark, kash@diku.dk

Abstract

Experiments help to understand human–computer interaction and to characterize the value of user interfaces. Yet, few intermediate guidelines exist on how to design, run, and report experiments. The present monograph presents such guidelines. We briefly argue why experiments are invaluable for advancing human–computer interaction beyond technical innovation. We then identify heuristics of doing good experiments, including how to build on existing work in devising hypotheses and selecting measures; how to craft challenging comparisons, rather than biased win–lose setups; how to design experiments so as to rule out alternative explanations; how to provide evidence for conclusions; and how to narrate findings. These heuristics are exemplified by excellent experiments in human–computer interaction.

Full text available at: http://dx.doi.org/10.1561/1100000043

Contents

Full text available at: http://dx.doi.org/10.1561/1100000043

Full text available at: http://dx.doi.org/10.1561/1100000043

1

Introduction

This work began as an attempt to answer a colleague's question. For some time I had insisted that we run experiments on a new interaction paradigm that we had been working on. My colleague had asked for papers that would convince him why we should do experiments at all. He also quickly asked for papers that explained how to do those experiments, seeing my expression of disbelief after the first question. I was unable, however, to give him entirely satisfactory references: this forms the background for the present work.

A fair number of papers describe how to do experiments in human–computer interaction (HCI). For instance, Landauer [86] gave a classic discussion of research methods in HCI, including valuable advice on statistical analysis and reporting. Blandford and colleagues [14] discussed how to plan, run, and report experiments in HCI, and presented an illustrative case study. Recently, Lazar et al. [90] published a book on research methods in HCI that included several chapters on designing and reporting experiments. Also, a number of papers review experimentation on topics closely related to HCI, including information retrieval [81], information visualization [21], and text editing [118]. More generally, a host of literature relevant to the design of

Full text available at: http://dx.doi.org/10.1561/1100000043

experiments exists in the field of psychology [95, 122], sociology [142], and ergonomics [34].

Why, then, another paper on experiments in HCI? First, the above papers focus little on the questions that arise even when you understand the basics of experimental logic, the distinction between independent and dependent variables, and the concerns in ensuring statistical conclusion validity. Second, many of the papers referenced above focus little on the specific difficulties arising from experimenting with interfaces and interactions. Third, while papers on specific topics are helpful, they de-emphasize that many areas of HCI face similar questions about why and how to do experiments.

We consider an experiment "a study in which an intervention is deliberately introduced to observe its effects" [127], p. 12. The intervention may be of a variety of kinds; in HCI it is often a technology, but could be kinds of training, user group, use situation, or task. We follow common practice by designating the intervention as a level of an independent variable, or as a treatment, or as a condition. The effects of the intervention are measured as dependent variables. In HCI they will often include measures of the usability of the technology. Hypotheses are statements that connect variation in independent variables to expectations about variation in the dependent variables. Another defining characteristic of experiments is that they attempt to deal with other factors besides the independent variable that influence the situation under study, and thus potentially affect the dependent variables [42]. This may happen, for instance, by controlling such factors, holding them constant, or distributing them randomly across levels of the independent variable. Finally, it is typical of experiments that the situation under study is created or initiated by the experimenter [42]. Figure 1.1 shows an outline of these components. Note that the above definition excludes the understanding implied in some common usages of the word experiment, including that of "trying something new" or "an innovative act or procedure".

The logic underlying experiments is tied to pioneering work in the renaissance, in particular by Galileo Galilei and Francis Bacon. Later, John Stuart Mill refined thinking about experiments by his Joint Method of Agreement and Difference. The key idea is that effects occur

Full text available at: http://dx.doi.org/10.1561/1100000043

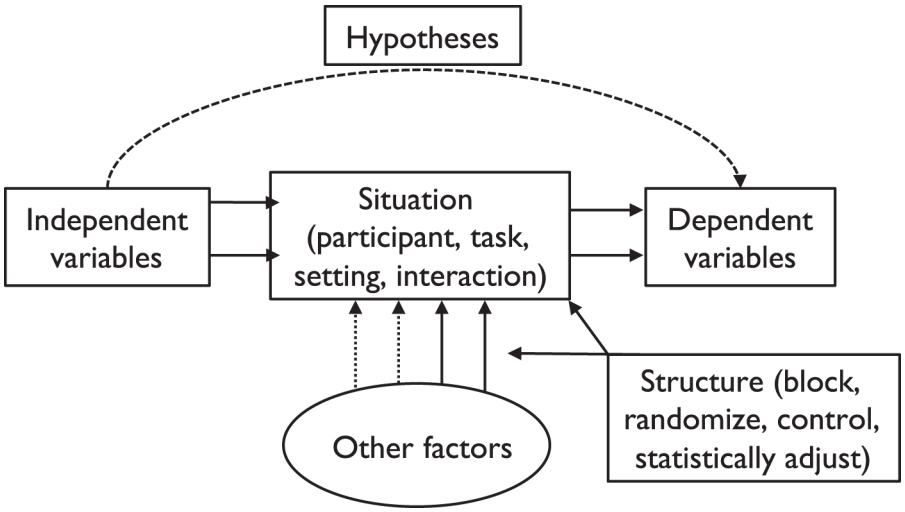

Fig. 1.1 Typical components of experiments in human–computer interaction.

with their presumed causes and that any difference between outcomes may be used to attribute causes; this idea is directly reflected in the above definition of experiment. Bunge [15] and Shadish et al. [127] further discuss the logic underlying experimentation and its historical development.

Full text available at: http://dx.doi.org/10.1561/1100000043

References

[1] R. Abelson, *Statistics as Principled Argument*. New York: Lawrence Erlbaum, 1995.

[2] J. Accot and S. Zhai, "Beyond Fitts' law: models for trajectory-based HCI tasks," in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pp. 295–302, New York, NY, USA: ACM, 1997.

[3] E. Adar, J. Teevan, and S. T. Dumais, "Large scale analysis of web revisitation patterns," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1197–1206, New York, NY, USA: ACM, 2008.

[4] R. Adcock and D. Collier, "Measurement Validity: A shared Standard for Qualitative and Measurement Validity: A shared Standard for Qualitative and Quantitative Research.," *American Political Science Review*, vol. 95, no. 3, pp. 529–546, 2001.

[5] D. L. Alonso, A. Rose, C. Plaisant, and K. L. Norman, "Viewing personal history records: A comparison of tabular format and graphical presentation using LifeLines," *Behaviour & Information Technology*, vol. 17, no. 5, pp. 249–262, 1998.

[6] C. A. Anderson, J. J. Lindsay, and B. J. Bushman, "Research in the Psychological Laboratory: Truth or Triviality?," *Current Directions in Psychological Science*, vol. 8, no. 1, pp. 3–9, 1999.

[7] J. Bailar and F. Mosteller, "Guidelines for statistical reporting in articles for medical journals," *Annals of Internal Medicine*, vol. 108, no. 2, pp. 266–73, 1998.

[8] R. Bakeman, "Behavioral observation and coding," in *Handbook of research methods in social and personality psychology*, (H. T. Reis and C. M. Judd, eds.), pp. 138–159, Cambridge, UK: Cambridge University Press, 2000.

Full text available at: http://dx.doi.org/10.1561/1100000043

[9] J. A. Bargas-Avila and K. Hornbæk, "Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2689–2698, New York, NY, USA: ACM, 2011.

[10] L. Barkhuus and J. A. Rode, "From Mice to Men - 24 Years of Evaluation in CHI," in *Paper presented at alt.chi*, 2007.

[11] J. Barzun, *Simple & Direct*. New York, NY: Harper Perennial, 2001.

[12] L. Berkowitz and E. Donnerstein, "External validity is more than skin deep: Some answers to criticisms of laboratory experiments.," *American psychologist*, vol. 37, no. 3, p. 245, 1982.

[13] N. Bevan, "Measuring usability as quality of use," *Software Quality Journal*, vol. 4, no. 2, pp. 115–130, 1995.

[14] A. Blandford, A. Cox, and P. Cairns, "Controlled experiments," in *Research methods for Human Computer Interaction*, (P. Cairns and A. L. Cox, eds.), Cambridge, UK: Cambridge University Press, 2008.

[15] M. Bunge, *Causality and modern science*. New York: Dover Publications, 3rd ed., 1979.

[16] M. M. Burnett, L. Beckwith, S. Wiedenbeck, S. D. Fleming, J. Cao, T. H. Park, V. Grigoreanu, and K. Rector, "Gender pluralism in problem-solving software," *Interacting with Computers*, vol. 23, no. 5, pp. 450 – 460, 2011.

[17] P. Cairns, "HCI... not as it should be: inferential statistics in HCI research," in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers*, pp. 195–201, Swinton, UK: British Computer Society, 2007.

[18] J. Campbell, "Labs, fields, and straw issues," in *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management*, (E. Locke, ed.), pp. 269–279, Lexington, MA: Lexington Books, 1986.

[19] J. Campbell, R. Daft, and C. Hulin, *What to study: Generating and developing research questions*. Beverly Hills, CA: Sage, 1982.

[20] S. K. Card, W. K. English, and B. Burr, "Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT," *Ergonomics*, vol. 21, no. 8, pp. 601–613, 1978.

[21] S. Carpendale, "Evaluating information visualizations," in *Information Visualization: Human-Centered Issues and Perspectives*, (A. Kerren, J. Stasko, J.-D. Fekete, and C. North, eds.), pp. 19–45, Berlin: Springer, 2008.

[22] T. Chamberlin, "The method of multiple working hypotheses," *Science*, vol. 15, no. 366, pp. 92–96, reprinted 1965, v. 148, p. 754-759., 1890.

[23] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 213–218, New York, NY, USA: ACM, 1988.

[24] A. Cockburn, C. Gutwin, and S. Greenberg, "A predictive model of menu performance," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 627–636, New York, NY, USA: ACM, 2007.

Full text available at: http://dx.doi.org/10.1561/1100000043

[25] A. Cockburn and B. McKenzie, "Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 203–210, New York, NY, USA: ACM, 2002.

[26] J. Cohen, *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum, 1988.

[27] J. Cohen, "A power primer," *Psychological bulletin*, vol. 112, no. 1, p. 155, 1992.

[28] H. Cooper, *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: Sage Publications, third ed., 1998.

[29] G. Cumming and S. Finch, "Inference by eye: Confidence intervals and how to read pictures of data," *American Psychologist*, vol. 60, pp. 170–180, 2005.

[30] M. Davis, "That's interesting," *Philosophy of the Social Sciences*, vol. 1, no. 2, p. 309, 1971.

[31] S. A. Douglas, A. E. Kirkpatrick, and I. S. MacKenzie, "Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 215–222, New York, NY, USA: ACM, 1999.

[32] P. Dragicevic, A. Bezerianos, W. Javed, N. Elmqvist, and J.-D. Fekete, "Temporal distortion for animated transitions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2009–2018, New York, NY, USA: ACM, 2011.

[33] M. Dunlop and M. Baillie, "Paper Rejected ($p > 0.05$): An Introduction to the Debate on Appropriateness of Null-Hypothesis Testing," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, vol. 1, no. 3, pp. 86–93, 2009.

[34] P. Edwards, F. Sainfort, T. Kongnakorn, and J. Jacko, "Methods of Evaluating Outcomes," in *Handbook of Human Factors and Ergonomics*, (G. Salvendy, ed.), pp. 1150–1187, Hoboken, NJ: Wiley, third ed., 2006.

[35] D. E. Egan, J. R. Remde, L. M. Gomez, T. K. Landauer, J. Eberhardt, and C. C. Lochbaum, "Formative design evaluation of superbook," *ACM Transactions on Information Systems*, vol. 7, no. 1, pp. 30–57, Jan 1989.

[36] D. Fanelli, "Negative results are disappearing from most disciplines and countries," *Scientometrics*, vol. 90, no. 3, pp. 891–904, 2012.

[37] G. Firebaugh, *Seven rules for social research*. Princeton, NJ: Princeton University Press, 2008.

[38] J. Fleiss, *Statistical methods for rates and proportions*. New York, NY: John Wiley & Sons, second ed., 1981.

[39] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-touch vs. mouse input for tabletop displays," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 647–656, New York, NY, USA: ACM, 2007.

[40] O. Frandsen-Thorlacius, K. Hornbæk, M. Hertzum, and T. Clemmensen, "Non-universal usability?: a survey of how usability is understood by Chinese and Danish users," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 41–50, New York, NY, USA: ACM, 2009.

Full text available at: http://dx.doi.org/10.1561/1100000043

[41] E. Frøkjær, M. Hertzum, and K. Hornbæk, "Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 345–352, New York, NY, USA: ACM, 2000.

[42] H. L. Fromkin and S. Streufert, "Laboratory Experimentation," in *Handbook of Industrial and Organizational Psychology*, (M. Dunnette, ed.), Chicago, IL: Rand-McNally, 1976.

[43] J. Garst, N. Kerr, S. Harris, and L. Sheppard, "Satisficing in hypothesis generation," *The American journal of psychology*, vol. 115, no. 4, pp. 475–500, 2002.

[44] V. Gawron, *Human performance, workload, and situational awareness measures handbook*. Boca Raton, FL: CRC, 2008.

[45] A. S. Gerber and D. P. Green, *Field experiments: Design, analysis, and interpretation*. New York, NY: WW Norton, 2012.

[46] R. Giner-Sorolla, "Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science," *Perspectives on Psychological Science*, vol. 7, no. 6, pp. 562–571, 2012.

[47] R. L. Glass, V. Ramesh, and I. Vessey, "An analysis of research in computing disciplines," *Communications of the ACM*, vol. 47, no. 6, pp. 89–94, Jun 2004.

[48] J. D. Gould, J. Conti, and T. Hovanyecz, "Composing letters with a simulated listening typewriter," *Communications of the ACM*, vol. 26, no. 4, pp. 295–308, Apr 1983.

[49] W. Gray and M. Salzman, "Damaged merchandise? A review of experiments that compare usability evaluation methods," *Human–Computer Interaction*, vol. 13, no. 3, pp. 203–261, 1998.

[50] S. Greenberg and B. Buxton, "Usability evaluation considered harmful (some of the time)," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 111–120, New York, NY, USA: ACM, 2008.

[51] A. Greenwald, "Within-subjects designs: To use or not to use?," *Psychological Bulletin*, vol. 83, no. 2, p. 314, 1976.

[52] A. Greenwald, A. Pratkanis, M. Leippe, and M. Baumgardner, "Under what conditions does theory obstruct research progress?," *Psychological review*, vol. 93, no. 2, p. 216, 1986.

[53] T. Grossman and R. Balakrishnan, "The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 281–290, New York, NY, USA: ACM, 2005.

[54] T. Grossman, G. Fitzmaurice, and R. Attar, "A survey of software learnability: metrics, methodologies and guidelines," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 649–658, New York, NY, USA: ACM, 2009.

[55] Z. Guan, S. Lee, E. Cuddihy, and J. Ramey, "The validity of the stimulated retrospective think-aloud method as measured by eye tracking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1253–1262, New York, NY, USA: ACM, 2006.

Full text available at: http://dx.doi.org/10.1561/1100000043

[56] C. Gutwin and S. Greenberg, "Effects of awareness support on groupware usability," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 511–518, New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1998.

[57] P. Harris, *Designing and reporting experiments in psychology*. Milton Keynes: Open University Press, 2008.

[58] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human mental workload*, (P. A. Hancock and N. Meshkati, eds.), pp. 139–183, North Holland: Elsevier, 1988.

[59] M. Hassenzahl, "Experience Design: Technology for all the right reasons," *Synthesis Lectures on Human-Centered Informatics*, vol. 3, no. 1, pp. 1–95, 2010.

[60] M. Hassenzahl, M. Burmester, and F. Koller, "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität," in *Mensch & computer*, pp. 187–196, 2003.

[61] M. Hassenzahl and A. Monk, "The inference of perceived usability from beauty," *Human–Computer Interaction*, vol. 25, no. 3, pp. 235–260, 2010.

[62] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world," *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 61–83, 2010.

[63] N. Henze, E. Rukzio, and S. Boll, "Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pp. 2659–2668, New York, NY, USA: ACM, 2012.

[64] M. Hertzum and E. Frøkjær, "Browsing and querying in online documentation: a study of user interfaces and the interaction process," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. 2, pp. 136–161, Jun 1996.

[65] D. M. Hilbert and D. F. Redmiles, "Extracting usability information from user interface events," *ACM Computing Surveys*, vol. 32, no. 4, pp. 384–421, Dec 2000.

[66] J. I. Hong and J. A. Landay, "WebQuilt: a framework for capturing and visualizing the web experience," in *Proceedings of the 10th international conference on World Wide Web*, pp. 717–724, New York, NY, USA: ACM, 2001.

[67] K. Hornbæk, "Current practice in measuring usability: Challenges to usability studies and research," *International Journal of Man-Machine Studies*, vol. 64, no. 2, pp. 79–102, 2006.

[68] K. Hornbæk, "Dogmas in the assessment of usability evaluation methods," *Behaviour & Information Technology*, vol. 29, no. 1, pp. 97–111, 2010.

[69] K. Hornbæk and E. Frøkjær, "Reading patterns and usability in visualizations of electronic documents," *ACM Transactions on Computer-Human Interaction*, vol. 10, no. 2, pp. 119–149, 2003.

[70] K. Hornbæk and E. L.-C. Law, "Meta-analysis of correlations among usability measures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 617–626, New York, NY, USA: ACM, 2007.

Full text available at: http://dx.doi.org/10.1561/1100000043

[71] W. Hwang and G. Salvendy, "Number of people required for usability evaluation: the 10±2 rule," *Communications of the ACM*, vol. 53, no. 5, pp. 130–133, 2010.

[72] ISO, *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*. Geneva, Switzerland: International Organization for Standardization, 1998.

[73] M. R. Jakobsen and K. Hornbæk, "Evaluating a fisheye view of source code," in *Proceedings of the 2006 Conference on Human Factors in Computing Systems*, pp. 377–386, New York, NY: ACM Press, 2006.

[74] Y. Jansen, P. Dragicevic, and J.-D. Fekete, "Evaluating the Efficiency of Physical Visualizations," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2593–2602, New York, NY, USA: ACM Press, 2013.

[75] S. Jarvenpaa, G. Dickson, and G. DeSanctis, "Methodological issues in experimental IS research: experiences and recommendations," *MIS quarterly*, vol. 9, no. 2, pp. 141–156, 1985.

[76] R. Jeffries, J. Miller, C. Wharton, and K. Uyeda, "User interface evaluation in the real world: a comparison of four techniques," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 119–124, New York, NY: ACM, ACM Press, 1991.

[77] B. John, "Evidence-based practice in human-computer interaction and evidence maps," *ACM SIGSOFT Software Engineering Notes*, vol. 30, no. 4, pp. 1–5, 2005.

[78] B. John and S. Marks, "Tracking the effectiveness of usability evaluation methods," *Behaviour & Information Technology*, vol. 16, no. 4-5, pp. 188–202, 1997.

[79] M. Kaptein and J. Robertson, "Rethinking statistical analysis methods for CHI," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pp. 1105–1114, New York, NY, USA: ACM, 2012.

[80] M. C. Kaptein, C. Nass, and P. Markopoulos, "Powerful and consistent analysis of likert-type ratingscales," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2391–2394, New York, NY, USA: ACM, 2010.

[81] D. Kelly, "Methods for evaluating interactive information retrieval systems with users," *Foundations and Trends in Information Retrieval*, vol. 3, no. 1-2, pp. 1–224, 2009.

[82] G. Keren, "Between-or within-subjects design: A methodological dilemma," in *A Handbook for Data Analysis in the Behaviorial Sciences*, (G. Keren and C. Lewis, eds.), Hillsdale, NJ: Lawrence Erlbaum, 1992.

[83] N. Kerr, "HARKing: Hypothesizing after the results are known," *Personality and Social Psychology Review*, vol. 2, no. 3, pp. 196–217, 1998.

[84] J. Kjeldskov, M. Skov, B. Als, and R. Høegh, "Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field," in *Mobile Human-Computer Interaction*, pp. 529–535, Springer, 2004.

Full text available at: http://dx.doi.org/10.1561/1100000043

[85] H. Lam and T. Munzner, "Increasing the utility of quantitative empirical studies for meta-analysis," in *Proceedings of the 2008 conference on BEyond time and errors: novel evaLuation methods for Information Visualization*, pp. 21–27, ACM, 2008.

[86] T. Landauer, "Research methods in human-computer interaction," in *Handbook of human-computer interaction*, (M. Helander, T. K. Landauer, and P. Prabhu, eds.), pp. 203–227, Amsterdam: Elsevier, second ed., 1997.

[87] P. Lang, "Behavioral treatment and bio-behavioral assessment: Computer applications," in *Technology in Mental Health Care Delivery Systems*, (J. Sidowski, H. Johnson, and T. Williams, eds.), Norwood, NJ: Ablex, 1980.

[88] K. Larson and M. Czerwinski, "Web page design: implications of memory, structure and scent for information retrieval," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 25–32, New York, NY, USA: ACM Press, 1998.

[89] T. Lavie and N. Tractinsky, "Assessing dimensions of perceived visual aesthetics of web sites," *International Journal of Human-Computer Studies*, vol. 60, no. 3, pp. 269–298, 2004.

[90] J. Lazar, J. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Chichester, UK: Wiley, 2010.

[91] J. Lewis, "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use," *International Journal of Human-Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.

[92] C. M. Lombardi and S. H. Hurlbert, "Misprescription and misuse of one-tailed tests," *Austral Ecology*, vol. 34, no. 4, pp. 447–468, 2009.

[93] R. MacCoun, "Biases in the interpretation and use of research results," *Annual review of psychology*, vol. 49, no. 1, pp. 259–287, 1998.

[94] S. Mahlke and M. Thüring, "Studying antecedents of emotional experiences in interactive contexts," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 915–918, New York, NY: ACM, ACM Press, 2007.

[95] D. Martin, *Doing psychology experiments*. Wadsworth Publishing Company, 2007.

[96] J. E. McGrath, "Methodology matters: Doing research in the behavioral and social sciences," in *Human-Computer Interaction: Toward the Year 2000*, (R.M.Baecker, J. Grudin, and W. A. Buxton, eds.), pp. 152–169, Los Altos, CA: Morgan Kaufmann Publishers, 1995.

[97] W. McGuire, "Creative hypothesis generating in psychology: Some useful heuristics," *Annual Review of Psychology*, vol. 48, no. 1, pp. 1–30, 1997.

[98] D. Meister, *Conceptual aspects of human factors*. Johns Hopkins University Press Baltimore, 1989.

[99] A. Monk, J. Carroll, S. Parker, and M. Blythe, "Why are mobile phones annoying?," *Behaviour & Information Technology*, vol. 23, no. 1, pp. 33–41, 2004.

[100] D. Mook, "In defense of external invalidity.," *American psychologist*, vol. 38, no. 4, pp. 379–387, 1983.

Full text available at: http://dx.doi.org/10.1561/1100000043

[101] M. Moshagen and M. Thielsch, "Facets of visual aesthetics," *International Journal of Human-Computer Studies*, vol. 68, no. 10, pp. 689–709, 2010.

[102] F. Mueller, S. Agamanolis, and R. Picard, "Exertion interfaces: sports over a distance for social bonding and fun," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 561–568, New York, NY, USA: ACM, 2003.

[103] T. Munzner, "Process and pitfalls in writing information visualization research papers," in *Information Visualization: Human-Centered Issues and Perspectives*, (A. Kerren, J. Stasko, J.-D. Fekete, and C. North, eds.), pp. 134–153, Berlin: Springer, 2008.

[104] C. Nass, Y. Moon, B. J. Fogg, B. Reeves, and C. Dryer, "Can computer personalities be human personalities?," in *Conference Companion on Human Factors in Computing Systems*, pp. 228–229, New York, NY, USA: ACM, 1995.

[105] W. Newman and A. Taylor, "Towards a methodology employing critical parameters to deliver performance improvements in interactive systems," in *Proceedings IFIP TC13 Seventh International Conference on Human-Computer Interaction*, pp. 605–612, Amsterdam: IOS press, 1999.

[106] W. Newman, "A preliminary analysis of the products of HCI research, using pro forma abstracts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 278–284, New York, NY, USA: ACM, 1994.

[107] R. Nickerson, "Null hypothesis significance testing: a review of an old and continuing controversy.," *Psychological methods*, vol. 5, no. 2, p. 241, 2000.

[108] J. Nielsen, *Usability engineering*. Boston, MA: AP Professional, 1993.

[109] K. O'Hara and A. Sellen, "A comparison of reading paper and on-line documents," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 335–342, New York, NY: ACM Press, 1997.

[110] J. S. Olson, G. M. Olson, M. Storrøsten, and M. Carter, "Groupwork close up: a comparison of the group design process with and without a simple group editor," *ACM Transactions on Information Systems*, vol. 11, no. 4, pp. 321–348, Oct 1993.

[111] A. Oulasvirta, "Fielding Usability Evaluation," *IEEE Pervasive Computing*, vol. 11, no. 4, pp. 60–67, 2012.

[112] A. Oulasvirta, M. Wahlström, and K. Anders Ericsson, "What does it mean to be good at using a mobile device? An investigation of three levels of experience and skill," *International Journal of Human-computer Studies*, vol. 69, no. 3, pp. 155–169, 2011.

[113] T. Paek, S. Dumais, and R. Logan, "WaveLens: a new view onto Internet search results," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 727–734, New York, NY, USA: ACM, 2004.

[114] G. Paolacci, J. Chandler, and P. Ipeirotis, "Running experiments on amazon mechanical turk," *Judgment and Decision Making*, vol. 5, no. 5, pp. 411–419, 2010.

[115] D. Parnas and P. Clements, "A rational design process: How and why to fake it," *IEEE Transactions on Software Engineering*, no. 2, pp. 251–257, 1986.

Full text available at: http://dx.doi.org/10.1561/1100000043

[116] A. Pirhonen, S. Brewster, and C. Holguin, "Gestural and audio metaphors as a means of control for mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 291–298, New York, NY, USA: ACM, 2002.

[117] J. Platt, "Strong inference," *Science*, vol. 146, no. 3642, pp. 347–353, 1964.

[118] T. L. Roberts and T. P. Moran, "The evaluation of text editors: methodology and empirical results.," *Communications of the ACM*, vol. 26, no. 4, pp. 265–283, Apr. 1983.

[119] C. Robson, *Real world research: A resource for social scientists and practitioner-researchers*. Oxford, UK: Blackwell, 2002.

[120] R. Rosenthal and R. Rosnow, *The volunteer subject*. New York, NY: John Wiley & Sons, 1975.

[121] R. Rosenthal and R. Rosnow, *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge, UK: Cambridge University Press, 1985.

[122] R. Rosenthal and R. Rosnow, *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill, 1991.

[123] P. M. Sanderson and C. Fisher, "Exploratory sequential data analysis: foundations," *Human-computer Interaction*, vol. 9, no. 4, pp. 251–317, Sep 1994.

[124] J. Sauro and J. R. Lewis, "Correlations among prototypical usability metrics: evidence for the construct of usability," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1609–1618, New York, NY, USA: ACM, 2009.

[125] A. Sears and B. Shneiderman, "Split menus: effectively using selection frequency to organize menus," *ACM Transactions on Computer-Human Interaction*, vol. 1, no. 1, pp. 27–51, Mar 1994.

[126] D. Sears, "College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature.," *Journal of Personality and Social Psychology*, vol. 51, no. 3, p. 515, 1986.

[127] W. R. Shadish, T. Cook, and D. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company, 2002.

[128] B. Shneiderman, *Designing the user interface*. Reading, MA: Addison-Wesley, 1987.

[129] B. Shneiderman and C. Plaisant, "Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies," in *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–7, ACM, 2006.

[130] J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, vol. 22, no. 11, pp. 1359–1366, 2011.

[131] D. Sjøberg, J. Hannay, O. Hansen, V. Kampenes, A. Karahasanovic, N. Liborg, and A. Rekdal, "A survey of controlled experiments in software engineering," *IEEE Transactions on Software Engineering*, vol. 31, no. 9, pp. 733–753, 2005.

Full text available at: http://dx.doi.org/10.1561/1100000043

[132] D. H. Sonnenwald, M. C. Whitton, and K. L. Maglaughlin, "Evaluating a scientific collaboratory: Results of a controlled experiment," *ACM Transactions on Computer-Human Interaction*, vol. 10, no. 2, pp. 150–176, Jun 2003.

[133] R. Soukoreff and I. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," *International Journal of Human-Computer Studies*, vol. 61, no. 6, pp. 751–789, 2004.

[134] W. Strunk and E. White, *The elements of style*. New York, NY: The Macmillan Company, 1959.

[135] W. Tichy, "Should computer scientists experiment more?," *Computer*, vol. 31, no. 5, pp. 32–40, 1998.

[136] E. Tufte, *Beautiful evidence*. Cheshire, CT: Graphics Press, 2006.

[137] T. Tullis and W. Albert, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. San Francisco, CA: Morgan Kaufmann, 2008.

[138] E. Webb, D. Campbell, R. Schwartz, L. Sechrest, and J. Grove, *Nonreactive measures in the social sciences*. Boston: Houghton Mifflin, 1981.

[139] R. Weber, "Editor's comments: the rhetoric of positivism versus interpretivism: a personal view," *MIS quarterly*, vol. 28, no. 1, pp. iii–xii, 2004.

[140] S. Whittaker, L. Terveen, and B. Nardi, "Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI," *Human–Computer Interaction*, vol. 15, no. 2-3, pp. 75–106, 2000.

[141] L. Wilkinson and T. T. F. on Statistical Inference, "Statistical methods in psychology journals: Guidelines and explanations.," *American psychologist*, vol. 54, no. 8, p. 594, 1999.

[142] D. Willer and H. Walker, *Building experiments: Testing social theory*. Stanford, CA: Stanford University Press, 2007.

[143] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 143–146, New York, NY, USA: ACM, 2011.

[144] J. Yi, Y. ah Kang, J. Stasko, and J. Jacko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1224–1231, 2007.

[145] R. Yin, *Case study research: Design and methods*. Thousand Oaks, CA: Sage Publications, third ed., 2003.

[146] M. Zelkowitz and D. Wallace, "Experimental models for validating technology," *Computer*, vol. 31, no. 5, pp. 23–31, 1998.

[147] S. Zhai, S. Conversy, M. Beaudouin-Lafon, and Y. Guiard, "Human on-line response to target expansion," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 177–184, New York, NY, USA: ACM, 2003.