

**Presentation Title:**

How many users is enough? Determining usability test sample size

**Presentation Format:**

Panel

**Targeted Audience:**

Intermediate to advanced

**Presentation strategy:**

Overview of methodology

**Topic Category:**

Usability method implementation

**Length of Presentation:**

90 minutes

**System, Product, or Project Focus:**

No specific system, product, or project orientation

**Keywords:**

Cost-justifying usability

User-centered design

Sample size estimation

**Learning objectives:**

Attendees will learn about the controversy over usability test sample size, including impact on ROI and test validity, and hear panelists' and other attendees' approaches to solving the sample size problem.

**How presentation will be conducted:**

Panelists will deliver position papers that include reviews of their own and others' articles on the topic of usability test sample size (45 minutes). During audience participation, attendees will be solicited for their own experiences and lessons learned from usability testing and sample size (45 minutes).

**Abstract (50 words):**

The topic of "how many users" is of great interest to usability specialists who need to balance project concerns over ROI and timelines with their own concerns about designing usable interfaces. The panel will review the current controversies over usability test sample size, test validity, and reliability.

**DETAILED DESCRIPTION OF CONTENT, WITH SESSION TIMELINE****Panelists' Presentation: Framing the Issue (45 minutes)**

Nielsen and Landauer (1993), Virzi (1992), and Lewis (1994) published influential articles on the topic of sample size in usability testing. In these articles, the authors presented a mathematical model for determining the sample size for usability tests. The authors presented empirical evidence for the models and made several important claims:


- a) Most usability problems are detected with 3 to 5 subjects
- b) Running additional subjects during the same test is unlikely reveal new information


- c) Most severe usability problems are detected by first few subjects

Virzi's stated goal of determining an appropriate number of test subjects was to improve ROI in product development by reducing the time and cost involved in product design. Nielsen and Landauer (1993), building on earlier work by Nielsen (1988; 1989) and Nielsen et al (1990), replicated and extended Virzi's (1992) original findings and reported case studies that supported their claims for needing only small samples for usability tests. The "small sample" claims and their impact on usability methodology have been popularized in Nielsen's (2000) widely read "useit.com" online column.

Since that time, a number of authors have challenged Virzi's and Nielsen's "small sample" findings on methodological and empirical grounds (Bailey, 2001; Spool & Schroeder, 2001; Woolrych & Cockton, 2001). Additionally, two large-scale experiments on usability test methods have been conducted that bear directly on Virzi and Nielsen's claims (Molich, et al, 1998; Molich et al, 1999).

The topic of "how many users" is of great interest to usability specialists who need to balance project concerns over ROI and timelines with their own concerns about designing usable interfaces. The goals of this panel discussion are to review the current controversies over usability test sample size and lead a discussion of the topic with the audience:


- 
- Examine the original "small sample" claims, including Nielsen's (1990), Nielsen and Landauer's (1993), Virzi's (1992), and Lewis' (1992, 1994) articles
  - Review the responses, including studies that deal with reliability of usability testing
  - Make suggestions and recommendations for follow up studies



Each panelist will present their perspective on the topic of usability sample size (15 minutes). The panelists will then lead discussion of the topic with attendees.



#### **Audience Participation (45 minutes)**

Some of the issues that will likely arise in the presentation of position papers and during audience participation are:

- 
- Role of usability testing in the user-centered design process
  - Importance of continued focus on ROI for usability testing: cost justification
  - Importance of making actionable recommendations
  - Role of and importance of experience of usability tester in detecting usability problems
  - Reliability and validity of usability results
  - Defining and testing with different user groups
  - Case studies: who has them? Attendees' lessons learned based on the impact of selecting a sample for usability testing.

## DESCRIPTION OF MATERIALS (HANDOUTS)

Attendees will receive a complete bibliography of articles that address the issue of usability test sample size as part of one of the panelist's position paper.

- 
- 
- Bailey, R. (2001). How reliable is usability performance testing? UI design update newsletter, Sept. 2001. <http://www.humanfactors.com/library/sep01.asp#bobbailey>
- Bias, R. (1992). Top 10 ways to muck up an interface project. IEEE Software, Nov., 95-96.
- Bias, R., & Mayhew, D. (Eds.) (1994). Cost-justifying usability. Academic Press: New York.
- Dumas, J., Sorce, J., & Virzi, R. (1995). Expert reviews: how many experts is enough? In Proceedings of the HFES 39<sup>th</sup> Annual Meeting.
- Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In Conference on Human Factors in Computing Systems: CHI 2000 Conference Proceedings (pp. 345-352). The Hague, the Netherlands: ACM Press.
- ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDT's) – Part 11: Guidance on usability. International Organization for Standardization.
- Kessner, M., Wood, J., Dillon, R. F. & West, R.L. (2001, April). On the reliability of usability testing. In J. Jacko and A. Sears (Eds.), Conference on Human Factors in Computing Systems: CHI 2001 Extended Abstracts. pp. 97-98. Seattle: ACM Press.
- Lewis, J. R. (1982). Testing small system customer set-up. In Proceedings of the Human Factors Society 26th Annual Meeting (pp. 718-720). Dayton, OH: Human Factors Society.
- Lewis, J. R. (1992). Sample sizes for usability studies: additional considerations. Technical Report 54.711 (pp. 1-18). Boca Raton, FL: International Business Machines.
- Lewis, J. R. (1993). Problem discovery in usability studies: A model based on the binomial probability formula. In Proceedings of the Fifth International Conference on Human-Computer Interaction (pp. 666-671). Orlando, FL: Elsevier.
- Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. Human Factors, 36, 368-378.
- Molich, R., Bevan, N., Curson, I., Butler, S., Kindlund, E., Miller, D., Kirakowski, J. (1998, June). Comparative evaluation of usability tests. In Proceedings of the Usability Professionals Association. Washington, DC: UPA.
- Molich, R., Thomsen, A.D., Karyukina, B., Schmidt, L., Ede, M., van Oel, W., & Arcuri, M. (1999, May). Comparative evaluation of usability tests. In Conference on Human Factors in Computing Systems: CHI 1999 Extended Abstracts (pp. 83-84). Pittsburgh, PA: ACM Press.
- Nielsen, J. (1988): Evaluating the thinking aloud technique for use by computer scientists, Proc. IFIP Working Group 8.1. International Workshop on Human Factors of Information Systems Analysis and Design (London, UK, 28-29 July 1988).
- Nielsen, J. (1989). Usability engineering at a discount. In Salvendy, G. and Smith, M.J. (Eds.), Designing and Using Human-Computer Interfaces and Knowledge Based Systems, Elsevier Science Publishers, Amsterdam, 394-401.
- Nielsen, J. (1993). Usability engineering. San Diego, CA: Academic Press.
- Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen and R. L. Mack (Eds.), Usability inspection methods (pp. 25-61). John Wiley: New York.
- Nielsen, J. (2000, March). Why you only need to test with 5 users. Alertbox, March 19, 2000. <http://www.useit.com/alertbox/20000319.html>
- Nielsen, J., Dray, S. M., Foley, J.D., Walsh, P., and Wright, P. (1990). Usability Engineering on a Budget, Proceedings of the IFIP INTERACT'90 conference (Cambridge, U.K., 27–31 August 1990).
- Nielsen, J. & Landauer, T. K. (1993, April). A mathematical model of the finding of usability problems. In Proceedings of ACM INTERCHI'93 Conference (pp. 206-213). Amsterdam, the Netherlands: ACM Press.
- Spool, J. M. (2001). Eight Is Not Enough -- UIETips 06/05/01. Electronic newsletter.

- Spool, J, & Schroeder, W. (2001, April). Test web sites: five users is nowhere near enough. In J. Jacko and A. Sears (Eds.), *Conference on Human Factors in Computing Systems: CHI 2001 Extended Abstracts* (pp. 285-286). Seattle: ACM Press.
- Virzi, R. A. (1992). Refining the test phase of usability evaluation: how many subjects is enough? *Human Factors*, 34(4), 457-468.
- Woolrych, A., & Cockton, G . (2001). Why and when five test users aren't enough. In J. Vanderdonckt, A. Blandford, and A. Derycke (Eds.), *Proceedings of IHM-HCI 2001 Conference, Vol. 2* (pp. 105-108). Toulouse, France: Cépadèus Éditions.