How many participants do you need for closed card sorting? A case study of an e-commerce website

Theodoros Dougalis | Aristotle University of Thessaloniki, Greece
Christos Katsanos | Aristotle University of Thessaloniki, Greece

CHIGreece 2025 | 24 - 26 September 2025

Introduction

- **Information Architecture (IA)**
  - Specifies how the content is structured and labeled[1,2]
  - Good IA is crucial for user navigation [1,2]
- **Card sorting**
  - A key method for designing and evaluating IA by understanding how users group content[1]
  - Two main variations: open (OCS) vs. closed sorting (CCS)
  - This paper focuses on closed card sorting
- **Close card sorting (CCS)**
  - Users sort content items into a <u>predefined</u> set of categories
  - Helps in evaluating or redesigning IAs

[1] Spencer, D. (2009). Card sorting: Designing usable categories. Rosenfeld Media.
[2] Rosenfeld, L., Morville, P., & Arango, J. (2015). Information architecture: For the Web and beyond (4 edition). O'Reilly Media.

Research motivation and goal

- **Research motivation**
  - A key question for any HCI method: How many participants are required [for reliable, cost-effective results]?
    - **Open card sorting**: 10-30 participants are required[1-3]
    - **Closed card sorting**: unknown, there is no study so far

- **Research goal**
  - What is the minimum number of participants needed to obtain reliable results from a closed card sort?

[1] Pechlevanoudis, C., Zilidis, G., & Katsanos, C. (2023). How Many Participants Do You Need for an Open Card Sort? A Case Study of E-commerce Websites. In *IFIP INTERACT 2023* (pp. 80-89). Cham: Springer Nature Switzerland.
[2] Tullis, T., & Wood, L. (2004). How many users are enough for a card-sorting study? UPA 2004
[3] Lantz, E., Keeley, J. W., Roberts, M. C., Medina-Mora, M. E., Sharan, P., & Reed, G. M. (2019). Card sorting data collection methodology: How many participants is most efficient? J. Classif.

Methodology (1/2)

- **Study context**
  - A closed card sort for a well-known Greek e-commerce website
- **Participants**
  - 191 participants (mostly students, mean age 21.2, 133 male, 58 female)
- **Categories & Cards**
  - Selected the cards based on Spencer's recommendations[1]
  - Technology-oriented section of the website's existing IA
    - **7 categories** (e.g., "Computers", "Audio", "Image", "Gadgets")
    - **45 cards** (e.g., "Activity trackers", "Cables", "Power banks", "Speakers")
- **Instruments & Procedure**
  - Remote asynchronous using online card sort tool (Card Sort, open source)
  - Google Forms for consent, instructions and demographics

[1] Spencer, D. (2009). Card sorting: Designing usable categories. Rosenfeld Media.

Methodology (2/2)

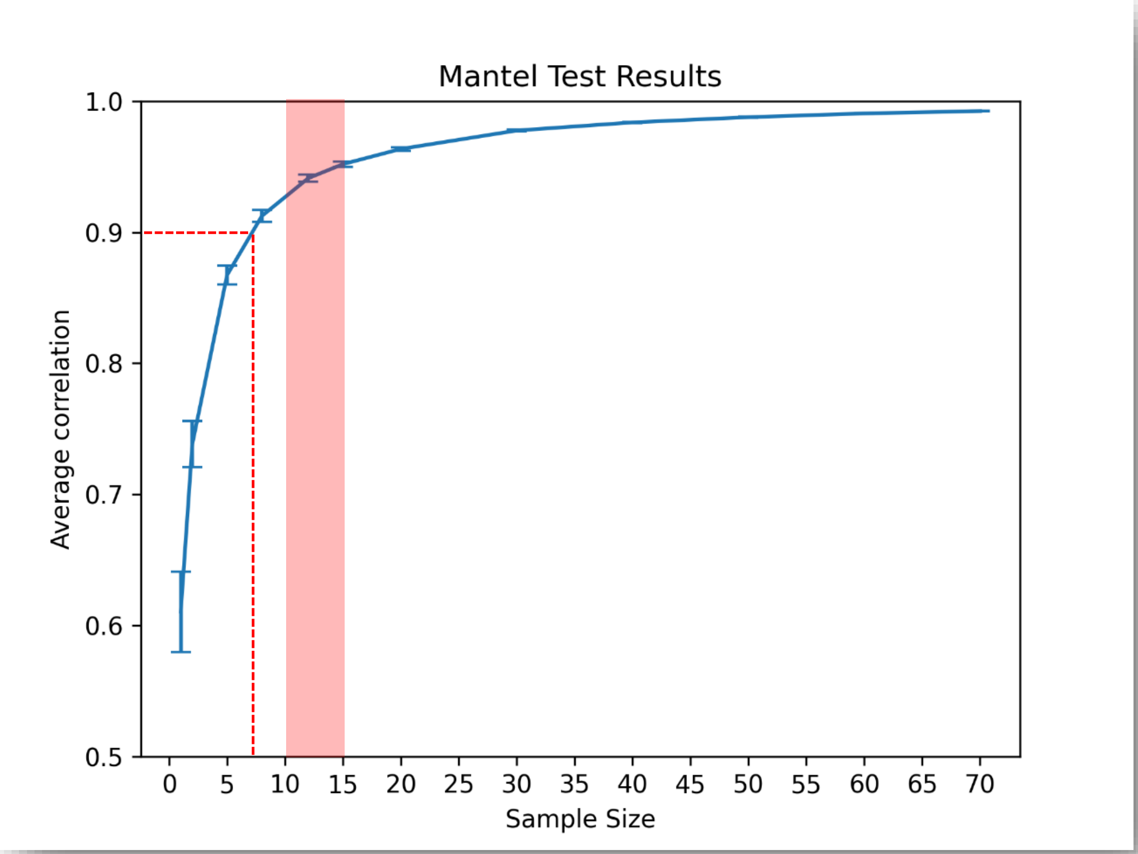
- **Data analysis methodology**
  - Resampling
    - Compare card sorts from N=191 (all) vs.N=samples of size M≤191
    - 100 random samples with size M=1, 2, 3 … 191
  - Analysis1: Distance matrices comparison (Mantel test)
    - Compare raw cards x cards groupings
    - Mantel correlation index (no assumptions that are violated for distance matrices)
  - Analysis2: Clusterings comparison (Elsim Score)
    - Compare cards x categories placements
    - Element-centric Clustering Similarity (Elsim) score: [0=dissimilar, 1=identical]
  - Custom-built data analysis tool written in Python

Results (1/2)

▪ Analysis1: Distance matrices comparison

- N=191 (All users)
- 100 random samples for each size M=1…N
- Error bars represent the 95% C.I.

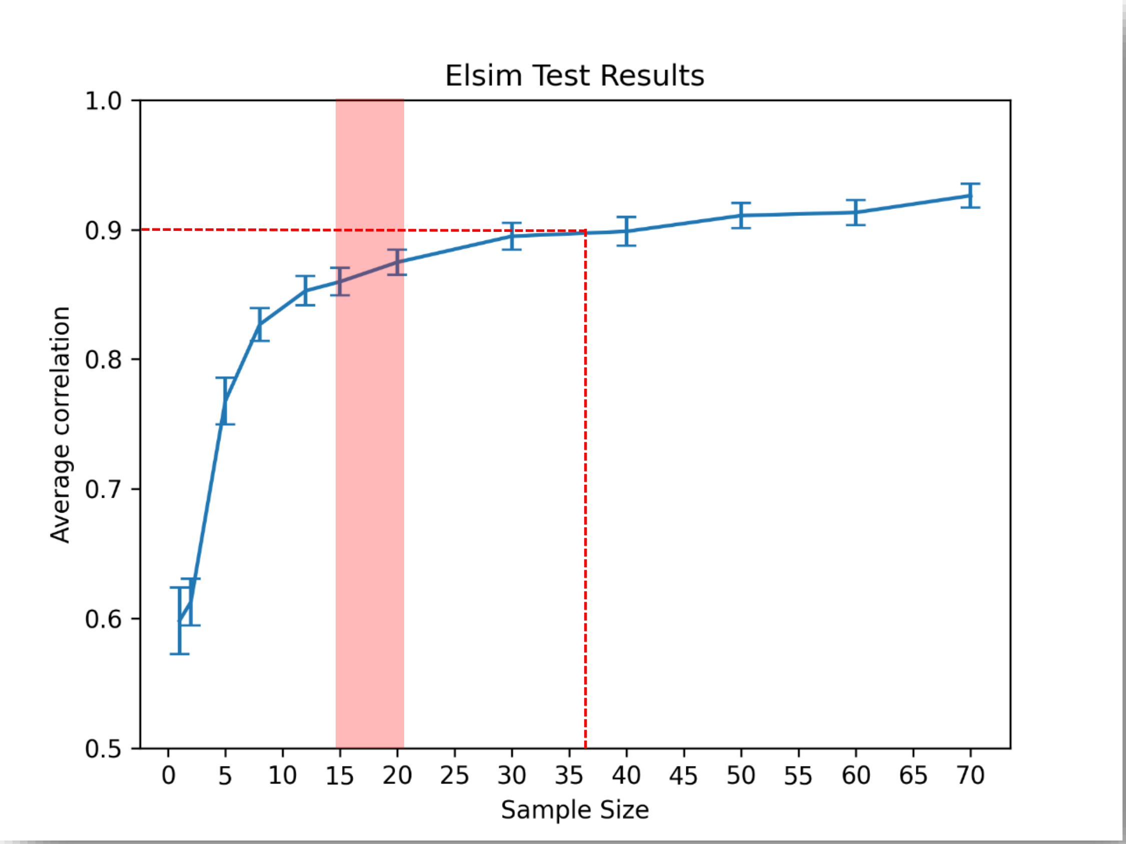
- Very little increase in Mantel r after 10-15 users
- r = 0.90[1] is achieved for 7 users



[1] Nielsen, J. (2004). Card Sorting: How many users to test. Alertbox. https://www.nngroup.com/articles/how-many-test-users/

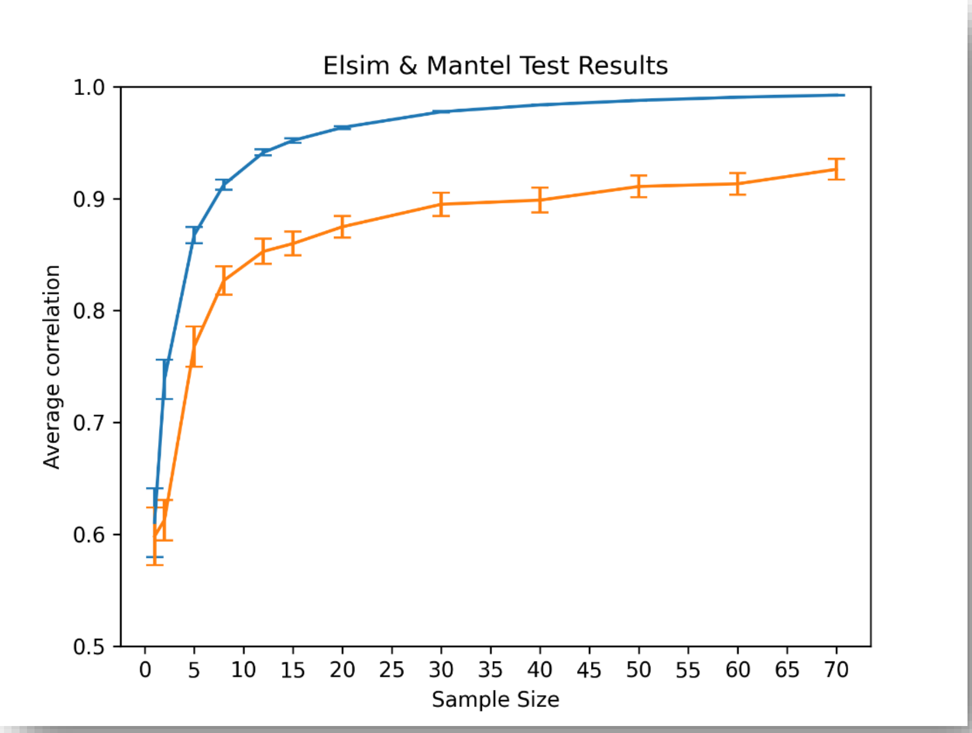
Results (2/2)

- **Analysis2: Clusterings comparison**
  - N=191 (All users)
  - 100 random samples for each size M=1…N
  - Error bars represent the 95% C.I.
  - Very little increase in Elsim score after 15 users
  - Elsim score = 0.90 is achieved for 35-40 users
  - Good compromise for cost efficiency: <u>12 to 21 users</u>; Elsim score = [0.853 to 0.882]



Discussion (1/2)

- **Mantel correlations were larger than Elsim similarity scores**
  - Mantel r = 0.90 for N=7 vs. Elsim score = 0.90 for N=35
  - Possible explanation: The Mantel test focuses on card co-occurrence, while the Elsim score focuses on the final groupings.
    - Two cards might be grouped together but in different categories, which only lowers the Elsim score.



Discussion (2/2)

- **The cost-benefit trade-off**
  - +5 users for +0.03 correlation improvement is waste of resources[1]
- **Our recommendation: N=8-12 participants for closed card sorts**
  - Raw data and cluster analysis results were already highly similar to the ones from all users; Mantel r = [0.91,0.94], Elsim score = [0.83,0.85]
  - For 12+ users, both metrics we used increased at a very low rate
- **Number of users for open vs. closed card sorting**
  - Our study's context closely aligns with that of Pechlevanoudis et al.[2] for OCS on e-commerce websites => N=15-20 users
  - CCS being more constrained than OCS (i.e., a more "closed" problem) produces reliable results with roughly half the number of users

[1] Nielsen, J. (2004). Card Sorting: How many users to test. Alertbox. https://www.nngroup.com/articles/how-many-test-users/
[2] Pechlevanoudis, C., Zilidis, G., & Katsanos, C. (2023). How Many Participants Do You Need for an Open Card Sort? A Case Study of E-commerce Websites. In *IFIP INTERACT 2023* (pp. 80-89). Cham: Springer Nature Switzerland.

Limitations & Future directions

- **Limitations**
  - Sample Demographics: The sample was gender-skewed (70% male) and primarily composed of university students
  - Single Tool & Context: The study used a single online tool and focused on one specific domain (e-commerce)
  - No Qualitative Data: The study did not collect qualitative feedback on participants' reasoning, which could provide additional insight

- **Future work**
  - Investigate the effect of gender, domain, and content complexity (number of cards/categories) on sample size
  - Extend the research question to other methods, such as hybrid card sorting

Summary & Questions

- **Summary**
  - We found that 8-12 users are required for the cost-effective collection of reliable data from closed card sorts
  - The study involved 191 participants sorting content from a real-world e-commerce site
  - Data analysis methodology that increases confidence on the findings (two types of analysis, 100 samples per size instead of 10, comparison for all possible sample sizes with a step of 1)
- **Questions?**
  - Shoot!
- **More questions and not enough time! No worries ☺**
  - Christos Katsanos (ckatsanos@csd.auth.gr)

Backup/Extra slides

Results

- **Descriptive statistics**

Size	Mantel r correlations				Elsim similarity scores			
N	Mean	SD	Min	Max	Mean	SD	Min	Max
1	0.610	0.153	0.181	0.859	0.598	0.128	0.329	0.887
3	0.788	0.067	0.616	0.902	0.690	0.092	0.429	0.887
5	0.867	0.037	0.778	0.937	0.768	0.090	0.580	1.000
8	**0.912**	**0.023**	**0.838**	**0.954**	**0.827**	**0.063**	**0.671**	**1.000**
12	**0.941**	**0.014**	**0.901**	**0.964**	**0.853**	**0.056**	**0.734**	**1.000**
15	0.952	0.011	0.923	0.971	0.860	0.053	0.692	1.000
20	0.964	0.007	0.937	0.978	0.875	0.049	0.771	0.945
30	0.978	0.004	0.966	0.986	0.895	0.053	0.780	1.000
40	0.984	0.003	0.974	0.989	0.899	0.055	0.780	1.000
50	0.988	0.002	0.981	0.992	0.911	0.049	0.817	1.000
60	0.991	0.001	0.987	0.993	0.913	0.049	0.799	1.000
70	0.992	0.001	0.989	0.995	0.926	0.046	0.799	1.000