

Exploring the Critical Decision Method for Usability Evaluation

Yiyao Li, Derin Delfin Borklu, Mojisolaoluwa Shakirat Idowu & Setia Hermawati

Human Factors Research Group, University of Nottingham

SUMMARY

Usability is central in the design and evaluation of interactive systems. There are different methods of usability studies, each offering some various advantages and challenges. The Critical Decision Method (CDM), widely used in cognitive task analysis, remains underexplored in usability contexts. This paper presents two empirical studies investigating CDM's applicability. Study 1 (n=16) examined the use of recall aids (screen recordings) with an entertainment recommender system. Results showed that recall aids only significantly enhanced recall accuracy in one out of two tasks but there was no statistically significant difference in usability insights, indicating that CDM can mostly be effectively implemented without additional memory support tools. Study 2 (n=16) compared CDM and Think Aloud (TA) in evaluating search interfaces. Findings revealed that CDM and TA are similar in their effectiveness in identifying usability problem topics; however, CDM provides more comprehensive cognitive insights. Two studies suggest that CDM has the potential to be a complementary or alternative method in usability evaluation.

KEYWORDS

Critical Decision Method, Usability Evaluation, Think Aloud

Introduction

Usability evaluation is a key part of Human-Computer Interaction (HCI), helping designers and developers identify potential interface issues before large-scale deployment. There are different methods of usability evaluation, each offering distinct advantages and challenges (Lazar et al., 2017). For instance, the commonly implemented Think Aloud (TA) method, which encourages participants to articulate their thinking processes during work with any system, displays user cognition and behaviour instantly (McDonald et al., 2012). However, the additional mental load to verbalise what goes on in one's mind when operating might compromise the richness of information during task execution (Honbaek, 2006; Bastien, 2010). Despite the widespread use of TA in usability testing, its limitations, such as increased cognitive load and potential distortion of natural user behaviour, highlight the need for alternative methods. Critical Decision Method (CDM), with its retrospective approach, may mitigate these issues by reducing interference during task execution while still capturing rich cognitive insights.

Understanding decision-making processes is critical for improving usability in complex systems. Existing methods often fail to capture the underlying reasoning behind user actions, which can limit the effectiveness of design improvements. CDM, traditionally used for Cognitive Task Analysis (CTA) of complex and critical incidents, has a retrospective approach where participants share thoughts on previous decision-making processes after task completion (Klein et al., 1989). It is defined as a methodology used to gather information by employing cognitive probes to analyse the strategies and rationales behind decisions made in non-routine situations (uncommon) and complex events (O'Hare et al., 1998; Plant and Stanton, 2013). CDM was selected as the primary

methodology for this research because it excels at eliciting non-observable mental processes, strategic shifts, and underlying rationales that standard usability methods might miss. While other CTA methods, such as Applied Cognitive Task Analysis (ACTA), were considered, they were deemed less practical for typical usability evaluation cycles. This is primarily because ACTA often requires multiple interaction sessions with participants to complete its various stages (Militello and Hutton, 1998). Recent studies have suggested that CDM may offer unique advantages for HCI (Hermawati et al., 2021; Cibulski et al., 2022). However, the use of CDM has been largely underexplored in the context of usability testing, especially for applications that involve common and simple events. This paper addresses these gaps through two empirical studies. Study 1 investigates the application of CDM for a usability evaluation of an existing application (an entertainment recommender system). Study 2 compares CDM with TA in usability evaluation of a prototype with medium fidelity (search bar). Together, these studies provide new evidence on methodological trade-offs, practical challenges, and opportunities of using CDM in usability evaluation.

Method

Study 1 involved 16 participants who were Netflix users. A between subject study was employed with participants split into two equal groups, with (Participants with Tools, PT) and without (Participants with No Tools, PNT) recall aid in the form of screen-recording as a memory assist tool during CDM interview. Each participant completed three tasks on Netflix: 1) Aimlessly scrolling through Netflix, with up to 5-minute time limit to make a final choice, 2) Selecting something they typically watch, up to 3 minutes time limit, and 3) Selecting something new or different, up to 3 minutes time limit. The duration of each CDM interview was recorded. For each participant, recall accuracy in each task during CDM interview was analysed by examining participant screen-recording videos to note every decision point they made and compared them with the number of steps recalled.

Study 2 also involved 16 participants. A between subject study was employed with participants split into two equal groups, TA and CDM groups. Participants performed two search tasks to locate specific articles. Participants were requested to search for academic articles by entering a relevant query, select the most appropriate suggestion, review the results, and ultimately identify the target article. Participants in the TA group were instructed to continuously verbalise their thoughts as they performed each task. In the case of the CDM group, participants' actions were recorded and used in the CDM interview following completion of the tasks. Themes of usability issues identified from TA and CDM were then extracted and compared.

The use of CDM probes during semi-structured interviews and subsequent analysis was a crucial qualitative feature in this study as shown in Table 1. CDM was used to analyse specific incidents through a guided retrospection process. According to Harrington et al. (2018) CDM allowed participants to explain the situation in their own terms, allowing both the analysis of the situation and exploration of possible options. Following an initial incident description, the CDM was utilised to delve into various aspects of the decision-making process. Subsequently, a detailed timeline of significant event descriptions was established.

Table 1: CDM probes and questions used in this study (adapted from Hoffman et al., 1998).

CDM Probes	Question
Goals	What were your specific goals at the various decision points? What were you trying to achieve?
Cues	What were you seeing or hearing that prompted your decision? What features were you looking at when making your decision?

Options	Were there any alternatives available other than your decision? Why were these alternatives not considered?
Expectations	What did you expect to find or achieve? How did your expectations influence your decisions?
Experience	What past experiences influenced this decision? What experience were you trying to replicate or avoid?
Knowledge-based	What information did you use in making this decision, and how was it obtained? What assumptions did you make based on this decision?
Situation Assessment	Did you use all the information available when formulating the decision? Was there any additional information you might have used to help form the decision?
Constraints	What limitations or restrictions did you have in making your decision?
Basis	What criteria or principle guided your decision? Can you suggest a rule based on your experience that could help others make the same decision successfully? Why or why not?
Level of Novelty	Does this situation fit a standard or typical browsing scenario for you? Was it something you were familiar with?
Time Pressure	How much time pressure did you feel when deciding what to watch? How long did it take you to make your decision?

Results

Study 1 revealed that the duration required to conduct CDM interviews with and without aid only showed significant difference in task 1 (aimlessly scrolling through Netflix to make a final choice). In task 1, participants with aid spent an average time of 10 minutes (SD = 5.41 minutes), whereas participants without aid spent an average time of 6.23 minutes (SD = 1.83 minutes); with t-test indicated significant difference ($t(14) = 1.8667$, $p\text{-value} = 0.0830$). In task 2, participants with aid spent an average time of 6.23 minutes (SD = 3.15 minutes), whereas participants without aid spent an average time of 4.29 minutes (SD = 2.13 minutes). In task 3, participants with aid spent an average time of 8.1 minutes (SD = 5.14 minutes), whereas participants without aid spent an average time of 6.53 minutes (SD = 2.19 minutes). The comparison of recall accuracy showed that there was no statistically significant difference between participants with and without recall aid for tasks that involved a higher degree of exploration (task 1 and task 3). However, in task 2, participants with aid gained higher recall accuracy than participants without recall aid ($t(7) = 3.547$, $p < 0.05$). The responses from the two groups were compared, by coding their responses and assigned them to themes, to evaluate the influence of recall aid on the richness of participants answers. A chi-square test showed no statistically significant difference between the groups regarding the number of unique themes of usability issues. These findings suggest that recall aids may not necessarily enhance usability insights, indicating that CDM can, in some circumstances, be effectively implemented without additional memory support tools especially for tasks that involve some a higher degree of exploration. This has practical implications for reducing resource requirements in usability studies. The group without recall aid provided more detailed and exploratory answers compared to the group with recall aid while the group with recall aid focused on the visible interface elements and immediate interactions to explain their decision points without providing in-depth reasons for their decisions in comparison with group without recall aid.

Study 2 showed that CDM and TA resulted in similar identification of usability issue themes. However, CDM yielded more complete cognitive insights, especially into why users made certain decisions and how they interacted with the search bar at a deeper level. It emerged from the study that while TA was effective in identifying immediate usability issues and user frustrations, CDM provided a better understanding of the underlying reasons behind such issues. The TA method

required participants to verbalise their thoughts during the task and often provided limited depth of insight into the underlying cognitive processes. The participants were liable to devote more time to describing what they did, rather than reflecting on why certain decisions were made. This resulted in a somewhat superficial level of understanding of the user's behaviour. On the other hand, the CDM method had an advantage in capturing deeper insights about the user's mental model and decision-making process. This stemmed from retrospective questioning: letting participants reflect on their actions after a task provides much richer, more detailed explanations of why they made certain decisions and how they navigated challenges. It captured not only the user's interaction with the specific interface but also the wider-scale cognitive processes, such as how prior experiences and expectations influenced their approach to the task. This makes CDM particularly valuable for gaining insights into how users think and make decisions in a broader sense than just the particular performance of the task. For example, CDM participants might explain why they favoured a given set of search strategies, adding in contextual information which might enable more focussed usability improvements. This level of detail is especially valuable when dealing with complex systems in which knowledge of the user's decision-making process is important. The ability of CDM to uncover deeper cognitive processes provides a unique advantage over TA, particularly for systems where understanding user reasoning is essential for design optimisation. This demonstrates CDM's potential as a complementary or alternative method in usability evaluation.

Discussion

Can CDM Be Used for Usability Study

The two studies presented in this paper explored the application of the CDM in usability testing, with Study 1 focusing on the role of recall aids and Study 2 comparing CDM to the TA method. Our findings suggest that the CDM, although originally developed for cognitive task analysis in high-stakes domains, can be successfully applied to usability evaluation of everyday interactive systems. In the entertainment recommender system, CDM uncovered how participants relied on visual cues, algorithmic trust, and prior experiences when making decisions. In the search usability context, CDM revealed strategies for balancing credibility, relevance, and efficiency. These findings highlight CDM's strength in going beyond the identification of surface-level usability problems to uncover why users behave as they do. This depth makes CDM a valuable complement to traditional usability methods. However, the method's feasibility depends heavily on research goals and available resources. CDM is more valuable for complex systems where the understanding of user cognition in terms of their thinking processes is as important as that about their interactions with the interface. CDM allows usability practitioners to capture how users transfer knowledge and strategies from one context to another and provides insight into how a user adapts to new systems or environments based on their prior experiences.

How to Use CDM for Usability Study

Study 1 highlighted the importance of recall aids, such as screen recordings, in mitigating memory limitations during CDM interviews. Recall aids improved accuracy in reconstructing decision points during routine-based task and enabled participants to articulate subtle reasoning that might otherwise be lost. However, at the same, recall aids also influenced participants to focus on the visible interface elements focus and consequently reduce the depth in usability insights and recommendations. Absence of recall aids, on the other hand, prompted participants to provide much more usability insights and recommendation at the cost of accurate recall. To address the question at hand, if we were to implement CDM in routine tasks today, which approach should we take: employing recall aids or not? The answer to this is that the decision should be based on whether the primary objective is to gather broader usability insights (favouring no recall aids) or to map detailed decision processes (favouring recall aids). In practice, a hybrid strategy that encourages recall to

gather more detailed experiential data while utilising assistance to establish a baseline of in-depth user behaviour may be effective. The mixed-method strategy might consider using tools first to jog one's memory, aiding recall. Then, the second stage can involve capturing overall impressions without using tools. Usability practitioners should be aware of the potential bias caused by aids and should strive to create an environment where users can freely express themselves without being limited only to what the screen recording or aids capture.

Comparison with TA Method

It emerged from the study that while TA was effective in identifying immediate usability issues and user frustrations, CDM provided a better understanding of the underlying reasons behind such issues. In other words, CDM yielded more complete cognitive insights, especially into why users made certain decisions and how they interacted with the search bar at a deeper level. However, as Cooke (1994) discussed, there is a trade-off between the depth of insight and cost/time investment; thus, the question remains whether such extra insights provided have enough value to justify the extended effort and cost. In relatively less complex usability contexts, where all that is needed is superficial problems to be found quickly, TA may be more practical and sufficient. This study suggests that while CDM provides deeper insights, the additional value must be carefully considered against the practicalities of the research context. For instance, in less complex systems and routine usability testing, the extra work that comes with CDM might not be necessary unless such a method becomes relevant for a deep understanding of user decision-making processes. From a researcher's perspective, CDM is undoubtedly more challenging to administer and analyse. The method requires a higher level of skill in conducting and interpreting the interviews, as well as in analysing the rich, qualitative data it produces. Agreeing with Boren and Ramey (2000) study that highlights TA's relative ease of use, if a researcher or a team is new to usability testing, TA would be easier to implement and analyse. TA's real-time data collection is more straightforward, requiring less post-processing, which makes it more accessible for novice researchers or in contexts where quick results are needed.

This study also has limitations. Both studies involved relatively small sample sizes of postgraduate students, which may limit generalisability to broader user populations. In addition, CDM interviews were conducted by experienced facilitators; the method's effectiveness may vary when applied by less experienced researchers. It is acknowledged that the tasks used in these studies represent relatively common user interactions. However, a more complex scenario might yield different results.

Conclusion

This paper investigated the potential of the CDM as a usability evaluation technique through two empirical studies. Study 1 examined the role of recall aids in an entertainment recommender system, demonstrating that screen recordings as recall aids significantly extended the CDM process and improved recall accuracy in only one of the three tasks, while not significantly affecting usability insights. Study 2 compared CDM with the TA method in search for usability tasks, showing that while TA was more time-efficient and effective at identifying immediate usability issues, CDM provided deeper and more nuanced insights into user cognition and decision-making. Future research should further explore ways to enhance the efficiency and effectiveness of CDM in usability evaluation, for instance by streamlining the CDM probes and validate its application in usability studies.

Key Takeaway

Our findings demonstrate that CDM not only identifies usability issues comparable to TA but also offers richer cognitive insights into user decision-making. This positions CDM as a valuable tool

for researchers and practitioners seeking to enhance usability evaluation, especially in contexts requiring an understanding of user reasoning.

References

- Bastien, J. M. C. (2010). Usability testing: A review of some methodological and technical aspects of the method. *International Journal of Medical Informatics*, 79(4), e18–e23.
<https://doi.org/10.1016/j.ijmedinf.2008.12.004>
- Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3), 261–278. <https://doi.org/10.1109/47.867942>
- Cibulski, L., Dimara, E., Hermawati, S., & Kohlhammer, J. (2022). Supporting domain characterization in visualization design studies with the Critical Decision Method. In *VISGuides 2022*. IEEE. <https://doi.org/10.1109/VISGUIDES57787.2022.00007>
- Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41(6), 801–849. <https://doi.org/10.1006/ijhc.1994.1083>
- Harrington, K., Brown, M., Pinchin, J., & Sharples, S. (2018). Decision-making within missing person search. *Cognition, Technology & Work*, 20(4), 665–680.
<https://doi.org/10.1007/s10111-018-0515-x>
- Hermawati, S., Cibulski, L., & Lawson, G. (2021). Towards Using the Critical Decision Method for Studying Visualisation-Based Decision-Making. *Ergonomics & Human Factors 2021*.
- Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the Critical Decision Method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40(2), 254–276. <https://doi.org/10.1518/001872098779480442>
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research, *International Journal of Human-Computer Studies*, 64(2), 79-102.
<https://doi.org/10.1016/j.ijhcs.2005.06.002>.
- Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 462–472.
<https://doi.org/10.1109/21.31053>
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction* (2nd ed.). Morgan Kaufmann.
- McDonald, S., Edwards, H. E., & Zhao, T. (2012). Exploring think-alouds in usability testing: An international survey. *IEEE Transactions on Professional Communication*, 55(1), 2–19.
<https://doi.org/10.1109/TPC.2011.2182569>
- Militello, L.G. and Hutton, R.J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, 41(11), 1618-1641.
<https://doi.org/10.1080/001401398186108>
- O'Hare, D., Wiggins, M., Williams, A., & Wong, W. (1998). Cognitive task analyses for decision centred design and training. *Ergonomics*, 41(11), 1698–1718.
<https://doi.org/10.1080/001401398186144>

Plant, K. L., & Stanton, N. A. (2013). What is on your mind? Using the Perceptual Cycle Model and Critical Decision Method to understand the decision-making process in the cockpit. *Ergonomics*, 56(8), 1232–1250. <https://doi.org/10.1080/00140139.2013.809480>