Evaluating the cognitive load of planning a ski tour

MICHELE CAZZARO, Università di Trento, Italy

Ski-tour planning is an information intensive activity that can determine the safety of the touring and where decision quality can be influenced by interface design. This study evaluates the usability and perceived cognitive load of skitourenguru.ch, an online ski-tour planning platform. The study asked five experienced ski-tourers to perform 3 different tasks on the website while using the think-aloud protocol, and assessing their workload with weighted NASA-TLX. Findings indicate that high-level information is conveyed effectively through color-coding, but route comparison tasks induce a considerable extraneous cognitive load due to a lack of a specific feature. Participants ended up frequently relying on memory and repeated navigation steps to compare routes. The findings of this study highlight some interface barriers that can increase cognitive effort in a critical decision making phase and suggest design priorities for minimizing the perceived workload in a safety critical setting.

1 Introduction

Backcountry and off-piste skiing is a popular sport in the Alps, although a potentially dangerous one. There is never zero risk of incurring in an avalanche and athletes rely on official bulletins, weather reports and first hand evaluation to avoid accidents. The life of ski-tourers depends on their evaluation so the stakes are very high. Furthermore, there is no universal bulletin, and an athlete must evaluate the risk he's exposed on each different tour.

1.1 Skitourenguru.ch

Skitourenguru.ch is an online planning tool tailored for ski-touring, grouping data from different organizational and open data sets. Data displayed are mainly trail information, maps, and avalanche bulletins. The platform allows users to discover new ski-tours, filter ski-tours and rank them by the calculated risk, display all the information about technical difficulties of a specific tour, display and links to sources for avalanche risk assessment information. The platform provides value by unifying the available information under a single application, and using way finding algorithms and mathematical modeling to adapt and reduce the avalanche risk exposure of hand drawn routes.

1.2 Research question

Ski-tourers must continuously evaluate the terrain they are about to face, starting from and building on the crucial planning stage. The efficacy of evaluation does not only depend on the accuracy of the data, but also in its presentation quality. This is why I set off to investigate the cognitive load of performing this operation on skitourenguru.ch. The research question I used to center this user study is:

> ***How effectively does skitourenguru.ch support ski-tourers in identifying and comparing low-risk routes, and what interface barriers hinder this process?***

The aim of this formative usability study is to identify specific usability challenges and cognitive bottlenecks. The findings will provide actionable insights for improving the interface, with the ultimate goal of supporting safer and more informed decision-making in the backcountry.

Author's Contact Information: Michele Cazzaro, michele.cazzaro@studenti.unitn.it, Università di Trento, Trento, Trentino Alto Adige, Italy.

2 Methods

2.1 Participants sampling

Following Nielsen's model of usability testing (Nielsen, 1993) I deemed appropriate to limit the number of participants to (N=5). Limiting the study to such a low number of participants was done also because of the qualitative nature of the study, and its formative purpose. This sample size should be acceptable to identify the most prevalent usability issues while allowing for deep, case-based analysis in a short-time frame. The participants were recruited with **convenience** and **snowball** sampling among the local ski touring community. To ensure representativeness of the sample, criteria of access were: **practical experience** with ski-touring and **theoretical knowledge** about avalanche risk assessment. The participation was voluntary and fully informed (Appendix A). Prior to the tasks, participants completed a background questionnaire detailing their demographics and domain experience. Domain experience metrics collected include:

- **Ski Touring Frequency**
- **Self-Rated Familiarity** on a 1–5 Likert scale with:
  - Avalanche risk concepts (e.g., danger levels, problem types, aspects)
  - Digital route planning tools (like Skitourenguru.ch, FATMAP, etc.)
  - Topographic map-reading skill

The participant's information was anonymized through the assignment of participants IDs.

2.2 Selected sample

The final sample was composed of 5 participants (3 male, 2 female; aged 24-57, M = 27.4 years), all meeting the inclusion criteria of having previous ski-touring experience and avalanche prevention knowledge. No dropouts occurred during the study, and all of the sessions were completed after gathering voluntary informed consent.

2.2.1 Demographics & expertise. Participants' occupations were (2) students, (1) engineer, (1) employee and (1) software developer, all holding at least a bachelor's degree. Their ski-touring frequency ranged from occasional (1-5 times/year; n = 1) to frequent (16+ times/year; n=1)

Expertise Domain	**Mean Rating (1-5 Likert)**
Avalanche risk prevention	3.4
Route planning tools	3.4
Map-reading	3.8
Digital tools	4.4

Table 1. Participants domain expertise

2.3 Procedure

All of the sessions followed a session script to ensure the standardization and order of tasks. The session were performed with the assist of the digital tool *Lookback* and were mixed between in-person (N = 3) and remote (N = 2). The estimated duration of the session was 35-45 mins, including identity confirmation, informed consent and background form, explanation of think-aloud protocol, tasks completion, brief interview and finally NASA-TLX questionnaire. The Think-aloud protocol was indeed the strongest methodological tool adopted in this study, as it allowed me to effortlessly understand the mental processes behind the participants decisions. The NASA-TLX questionnaire is administered only once towards the end of the session to avoid continuously interrupting the users; this choice is further justified by the lack of need for a task specific cognitive load index. All sessions were recorded (screen, audio, and timestamps) to allow for later verification and systematic data coding.

2.4 Tasks

Participants executed a series of pre-defined route planning tasks designed to force interaction with the avalanche risk data in different ways. For the purposes of planning within the study, the website was set on a demo state containing historical data from a selected day during the past season.

2.4.1 Task 1: Exploratory route finding under constraints.

- **Goal:** Find a suitable ski tour that demonstrates low avalanche risk, has an ascent between 800 m and 1000 m, and a starting point within 150 km from Trento.
- **Cognitive focus:** Assess how users initially explore the interface, discover and use filters, and interpret "low risk" from the available visualizations (e.g., color overlays, icons).
- **Success criteria:** Selected route meets all three constraints.
- **Data logged:** Time to completion, filters applied, and participant comments.

2.4.2 Task 2: Comparative risk assessment.

- **Goal:** Compare two pre-selected, neighboring routes with different risk profiles. Determine which one is safer and justify the choice.
- **Cognitive focus:** Evaluate how users directly compare risk visualizations, interpret the color scale and symbols, and articulate their risk reasoning. This task specifically probes decision-making heuristics.
- **Success criteria:** The choice is logically justified based on the observed risk information, even if the "objectively" safer route is debatable.
- **Data logged:** Time to decision, and hesitation.

2.4.3 Task 3: GPX download (usability benchmark).

- **Goal:** Download the GPX file for the route selected in Task 2.
- **Cognitive focus:** Identify basic usability issues in the workflow that add unnecessary cognitive load or frustration, detracting from the primary planning task.
- **Success criteria:** GPX file is successfully located and downloaded.
- **Data logged:** Total time.

This task, although not a risk assessment task, is a typical task performed in preparation of an outdoor activity. Its purpose is to test usability frictions in the whole workflow.

2.5 Data collection and analysis plan

2.5.1 Data collection. Participants are set to complete 3 tasks each during a supervised usability study on skitourenguru.ch while vocalizing their thoughts. During the study, participants are required to fill in a custom demographics and background form, as well as a post-task weighted NASA-TLX form to capture their workload on a scale going from 1-10 across 6 different subscales. Completion metrics and timings are derived from the recordings of the session, together with intresting reflections participants have.

2.5.2 Data analysis. For the quantitative parts I computed means and SDs as well as the weighted totals from the NASA-TLX data. The task load index is collected only once towards the end of the session for 2 separate reasons:

(1) the study aimed at capturing the overall cognitive load of planning a ski-tour, and in real world conditions route discovery, comparison and GPX download are experienced as a continuous workflow.
(2) repeated administration of the long NASA-TLX after each task would have introduced frequent interruptions and increased participant's fatigue, potentially poisoning the results.

The qualitative data was reviewed to identify frustration instances and thematically group them.

2.5.3 Validity. The reliability of the findings is enhanced through the usage of triangulation (think-aloud + NASA-TLX), to mitigate the small sample size. The vocalizations helped cross-validate the logged metrics.

3 Results

3.1 Quantitative results

The data extracted from the NASA-TLX revealed an overall low to moderate workload (M=4.7, SD=1.35) during the completion of the tasks. **Mental** demand was the highest dimension of cognitive load ($M = 1.59$), followed closely by Performance ($M = 1.36$). The lowest dimension reported was unsurprisingly the **Physical** demand ($M = 0.03$). Higher perceived workload co-occurred with a greater number of navigation errors during GPX download, an observation treated as exploratory given the study's scale and design.

NASA-TLX Subscale	Mean (s)	StDev (s)
Mental	1.59	0.52
Performance	1.36	0.38
Effort	1.10	0.89
Frustration	0.34	0.46
Temporal	0.26	0.22
Physical	0.03	0.06
Total workload	**4.69**	**1.35**

Table 2. Mean task loads

Task	Mean (s)	StDev (s)
1	202	88,1
2	290	21,2
3	94,4	103,1

Table 3. Average time on task

Due to the small scale of the study these results cannot be considered inferential: these are considered informative of the collected qualitative data.

3.2 Qualitative Results

The analysis of participants' vocalized thoughts during the sessions produced the identification of four recurring themes across the five sessions, distributed over 28 total occurrences. One theme is tied to recurrent positive feedback.

3.2.1 Overview of themes.

(1) Menu confusion
(2) Issues with the map
(3) Comparison difficulty
(4) Color coding

3.2.2 Menu confusion. Participants vocalized frustration about the right-side menu of the website. This issue occurred in all five sessions. Some participants lamented a lack of structure in the menu, pointing to the absence of an evident navigation tree. Additionally, two participants were frustrated by the donation request banner occupying a large portion of the menu and distracting them from the tasks.

> *"I still only see a lot of useless advertising. Let's see menu, media, manual, information. No, all useless info, whatever"*

> *"As for here now on the menu, the avalanche risk is the only thing communicated well. For the rest there's the damn IBAN, an orange banner here, horrible, very annoying, which is a bit distracting."*

On another occasion, a participant got stuck in the modal interaction used to draw custom routes, confusing the icon to close the menu with the icon used to exit the route drawing mode.

> *"Help, I froze, I'm reloading the page because I'm an idiot and why did I make the little drawing?"*

In general, participants were surprised by the lack of usage of standardized design patterns in the menu, and expected the presence of a specific option to compare different tours.

> *"Basically, I had to think about it a bit more than I imagined and also a bit the use of all the functions, in the sense that nothing is too clear,"*

> *"Where is 'Compare tours' in the menu? [...] Whatever, let's do it this way."*

This was a navigation issue for the users, which impacted **all three Tasks.**

3.2.3 Issues with the map. Another recurring pain point during task completion was the struggle to identify elements on the map. This occurred for two main reasons:

- participants looking for known routes that were not available in the system
- the overlapping of routes obstructed the selection of a specific tour

> *"I don't know why here [. . .] I see Monte Sole, but if I click on it, it gives me Monte Villar, which is just a bit further over [. . .] Oh the routes completely overlap here."*

> *"Then I think it does not have the route in the system."*

This issue had an impact on **Tasks 1 & 2**.

3.2.4 Comparison difficulty. Almost all participants ($N = 4$) faced difficulties when trying to compare two similar routes in **Task 2**. Not finding a dedicated option to compare two routes, participants resorted to switching back and forth between routes on the map.

> *"Fantastic so to compare I have to stay on the same map I have to remember which mountain it is?!"*

> *"Where is 'Compare tours' in the menu?"*

This confusion and the lack of a dedicated function led to increased time on task for **Task 2**. Moreover, having to keep in mind a large amount of information while browsing routes may have increased participants' cognitive load.

3.2.5 Color coding. Participants liked the use of color coding and found it very useful in conveying high-level information on many occasions. It supported faster filtering of routes and the overall evaluation process.

> *"[For] the safety, you just look at the color, so it's very easy."*

> *"The color of the map helps you understand, the reddish accentuates it even more"*

> *"It shows you the steepest points, the colors are quite intuitive"*

4 Discussion

The study examined Skitourenguru's ability to support ski-tourers while choosing a tour, simulating what would happen in a natural environment, focusing on participant's perceived workload. The results convey that while the platform conveys high-level information about avalanche risk effectively through color-coding and visual overlays, other design choices introduce extraneous cognitive load. During the comparison between 2 different routes, participants had to rely on memory, repeated navigation or other ad-hoc strategies while making safety-critical decisions.

Across tasks, the most significant bottleneck emerged in fact in Task 2. Participants consistently struggled to compare routes due to the lack of a dedicate compare feature. Users resorted to memorizing risk attributes (slope, avalanche risk, exposure, color coding) in their working memory. Referring to the Cognitive Load Theory (Sweller, 1988), we can see how the interface in this specific case fails to minimize *extraneous load*, by not providing a sufficient perceptual support to the user during the task. This is supported by the numerous comments participants vocalized during the tasks, and by the consistency of high timing to complete the task. In this study, extraneous cognitive load is intended as effort that is unrelated to understanding avalanche risk itself.

Navigation and menu structure issues were another source of extraneous load affecting all tasks. A difficulty in understanding the right-side menu hierarchy and navigation, added to the distracting donation banner prompted frustration and was a source of confusion. In a safety-critical situations as in ski tour planning this isn't a mere distraction since it can interfere with users' focus on the task and lead to a slower and more imprecise access to information. Similarly, users that entered "Draw your route" mode were lost, having difficulty in quickly determining how to exit this mode. This suggests an insufficient visibility of this mode being active and the lack of usage of a standardized pattern on how to confirm or exit the mode lead to a temporary halt in tasks progress.

In contrast, the website use of color coding and map visualization were on multiple occasions signaled as strengths and appreciated by participants. Users found the use of color intuitive to understand and helpful in quickly visually filtering options. This suggests that the website adequately visually supports information scanning in a route exploration scenario.

The overall NASA-TLX scores indicate a low to medium workload. This does not directly translate into a measure of good usability: since the Task Load Index was taken after the completion of all the 3 tasks, it could also mean an unevenly distributed workload across tasks. Supporting evidence is found in the qualitative data as participants were more vocal about frustrations during the comparison task, which could be tied to a spike in mental demand. Given the centrality of comparing options when safely choosing a ski-tour these spikes can be considered problematic.

4.1 Limitations

This study has several limitations. First of all, it is not to be considered highly generalizable due to the small sample size ($N = 5$) obtained through convenience sampling. This created an uneven population of young, highly educated and digitally skilled individuals, which do not accurately represent the population of alpine ski tourers. As a result, the identified usability problems likely underestimate the difficulties of less digitally experienced or novice ski tourers.

Additionally, the study was performed during summer, when official up-to-date information is not available due to the lack of snow. The use of historical data in the platform's demonstrative setting may not fully capture the time-pressure of planning a tour during the ski season, as well as failing to capture actual real-world conditions.

Finally, measuring the workload globally rather than per task, while a smart logistic choice, limits the granularity of cognitive load distribution.

4.2 Design Implications

As a result of the acknowledgements above, there are ways to improve the cognitive load distribution and possibly reduce it by taking action in the design of the application. Adding a dedicated feature to compare side-by-side two or more routes would immediately ease pressure of actions resembling Task 2 in a natural setting. This could result in a reduction of reliance on working memory for comparisons, ultimately freeing up precious mental resources of the users.

It's understandable for a non-profit organization to aggressively seek funding between the users of their services. However, if the goal is to prioritize safe planning of winter activities in the backcountry, deriving from the findings of this study it is advisable to reduce the distraction caused by the donation banner and other elements not related directly to ski-tour planning. An actionable example would be swapping the current donation banner with a dismissible one, or displaying it along the initial popup to accept service terms.

Finally, clarifying that the only way to exit "Draw routes" mode is to trash the current drawing by clicking on the corresponding icon could be helpful in reducing frustration or the need for users to reload the page because of being stuck. Another option is to use a different icon conveying a different meaning, like suggesting the completion of the task.

5 References

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. Qualitative health research, 3(4), 430-441.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139-183). North-Holland.

Lookback.io. 2026. Lookback: User Research Platform. https://www.lookback.io/.

Nielsen, J. 2000. Why You Only Need to Test with 5 Users. Nielsen Norman Group. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Nielsen, J., & Landauer, T. K. (1993, May). A mathematical model of the finding of usability problems. In Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems (pp. 206-213). https://doi.org/10.1145/169059.169166

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). The think aloud method: a practical approach to modelling cognitive processes. London: AcademicPress, 11(6).

A Appendix A - Supplementary materials

All the documents listed in this appendix are available as an attachment to this PDF.

A.1 Consent form

All participants provided informed consent prior to participation. The consent form detailed the study purpose, procedures, data handling, anonymization, and participants' rights, including the right to withdraw at any time without consequence.

A.2 Raw NASA TLX results

Workload ratings collected during the usability study are provided as a separate anonymized CSV file. The dataset includes weighted NASA-TLX subscale ratings. Each row corresponds to one participant. These data are provided for transparency and replicability and were used exclusively for descriptive analysis in this study.

A.3 Raw Demographics

Participant background information was collected via a pre-study questionnaire. The raw, anonymized responses are provided as a CSV file (. The dataset includes self-reported demographic information (e.g., age range, gender), skiing and ski-touring experience, and familiarity with digital mapping tools. No directly identifying information was collected or stored.