doi: 10.36689/uhk/hed/2026-01-027 HED Hradec Economic Days

Evaluating the Usability of Interactive News Maps Using Eye-Tracking

Tomas VANICEK*, Jonas KOSTAL and Stanislav POPELKA

Palacký University Olomouc, Olomouc, the Czech Republic; tomas.vanicek@upol.cz; jonas.kostal01@upol.cz; stanislav.popelka@upol.cz

* Corresponding author: tomas.vanicek@upol.cz

Abstract: Interactive maps are nowadays routine in data journalism, but their usability depends on whether readers notice and understand interactive controls within busy narrative webpages. This paper reports a formative laboratory evaluation of interactive news maps using screen-based eye-tracking and talk-aloud (concurrent verbalisation). Forty-four participants (22 domain experts; 22 lay readers) completed eight information-extraction tasks across five journalistic articles under standardised time limits. Analysis combined qualitative behavioural diagnosis with quantitative indicators of effectiveness (task success), efficiency (time-on-task), and error-related patterns, supplemented by a composite performance score. Mean task success was 57.67%, with marked task-to-task differences (highest: 79.55% and 77.27%; lowest: 20.45% and 34.09%). The least successful tasks also took the longest, a pattern consistent with breakdowns associated with interaction demands. An analysis of Areas of Interest (AOIs) on article screenshots further shows that, during free exploration, participants allocated substantially more visual attention to article text than to graphical content. Overall, the findings highlight interaction discoverability and transient feedback as recurring bottlenecks, motivating practical design recommendations for making interactive news maps easier to notice, interpret, and use.

Keywords: interactive news maps; data journalism; usability evaluation; eye-tracking; interaction discoverability; formative evaluation

JEL Classification: D83; L86

1. Introduction

Interactive cartographic visualisations are a routine component of contemporary digital communication (Molina Rodríguez-Navas et al., 2021). Beyond “classic” web mapping, interactive maps increasingly appear as embedded artefacts within broader narrative environments. This is especially visible in online journalism, where maps are combined with text, images, and other interactive graphics to support data-driven storytelling and reader exploration (Rafeeq, 2025). Interactive narrative visualisation has been described as an emerging class of presentations that blends a guided story structure with opportunities for user-driven interaction. This pattern is particularly prevalent in online news production (Segel & Heer, 2010). In this setting, maps are not merely decorative. They can serve as

analytical interfaces through which readers interrogate spatial patterns, compare places, and draw conclusions that may influence their attitudes and decisions.

From a data-journalism perspective, interactivity is often viewed as a marker of quality and innovation. Empirical evidence, however, suggests a more nuanced picture. Content-analytic studies indicate that many daily data-driven stories still employ limited interactivity (Karypidou & Veglis, 2019; Zhang et al., 2023). When interactivity is present, maps tend to be among the most interactive visualisation types compared with other chart forms (Stalph, 2018). This creates a design tension. Editorial teams deploy interactive maps to communicate complex socio-demographic phenomena. Yet, the surrounding web context (scrolling articles, competing graphics, advertisements, cookie banners, and heterogeneous layouts) can interfere with map use, reduce the discoverability of interactive controls, and fragment attention. In practice, even well-intended interactive features may become "invisible" (Blascheck et al., 2019). This can happen when users do not notice them, fail to infer interactivity from the visual design, or do not understand how interaction changes the map's message. Recent practitioner-oriented research frames this as a map maker's dilemma: interactive maps promise richer public understanding, but real-world production is constrained by barriers spanning data acquisition and quality, technical implementation, user communication, and organisational context (Zbiejczuk Suchá et al., 2026). This perspective underlines that usability problems in journalistic mapping rarely stem from a single interface flaw. They often reflect systemic constraints that limit iteration, testing, and the clarity of interaction cues.

These challenges motivate systematic usability evaluation. Usability is commonly framed as the extent to which specified users can achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO, 2018). This definition is particularly relevant for interactive news maps because the context of use is close to everyday news consumption rather than to a controlled analysis session. Readers typically skim, scroll, and switch attention between text, graphics, and interface controls. Moreover, journalistic maps often address politically and socially salient topics where misinterpretation can have downstream consequences. Examples include misunderstanding relative versus absolute values, misreading classification, overlooking uncertainty, or conflating correlation with causation. Evaluating usability in this domain is therefore not only a matter of interface convenience. It is also essential for ensuring reliable public communication of spatial data.

In cartography and geovisualisation research, controlled user studies have long been used to assess map design and map use. Eye-tracking, in particular, has been used extensively to study how users allocate visual attention while reading and interacting with maps (Fairbairn & Hepburn, 2023). Overviews of eye-tracking applications in cartography emphasise its value for investigating perceptual and cognitive processes in map use. They also highlight methodological challenges and the need for careful experimental design (Krassanakis & Cybulski, 2021). Eye-tracking captures eye movements (e.g., fixations and saccades). These events are commonly used as proxies for visual attention and allow researchers to link user performance and errors to specific interface elements, spatial regions, and interaction sequences. Practical guidance on transforming raw gaze samples into

analysable events and measures (e.g., using Areas of Interest (AOIs), scanpaths, and aggregated metrics) is well documented in the methodological literature (Holmqvist et al., 2011; Popelka, 2018). Importantly, recent eye-tracking work (Popelka et al., 2025) on journalistic election visualisations shows how attention data can reveal subtle breakdowns. Users may fixate on a control that enables switching to an alternative map view, yet fail to recognise its purpose, resulting in underuse of the intended affordance and potential misinterpretation (Popelka et al., 2025). Such findings support the need to evaluate not only whether users view key elements, but also whether they correctly interpret and act upon them.

In interactive map interfaces, usability evaluation is critical because interactions change the information available to the user. Early eye-tracking work on interactive map interface design proposed combining gaze analysis with traditional usability metrics to understand user strategies better and identify design weaknesses that can remain hidden in log files or questionnaires (Çöltekin et al., 2009). A persistent methodological complication is that interactive exploration produces non-identical stimuli across participants. Pan/zoom, layer switching, pop-ups, and responsive user interface states can vary substantially between users and across time (Vanicek et al., 2024). This reduces straightforward comparability and encourages analysis strategies that can accommodate dynamic interface states, such as dynamic AOIs, event-based segmentation, or hybrid approaches combining gaze with interaction logging.

Because gaze data indicate where users look but not necessarily why, eye-tracking is often complemented by verbal methods (Vanicek & Popelka, 2023). Protocol-based verbalisation techniques treat spoken reports as data that can reveal information accessed in working memory during task performance, when elicitation is appropriately designed and does not unduly alter the cognitive process being studied (Ericsson, 2006). In usability and human–computer interaction practice, talk-aloud (concurrent verbalisation) is often used to expose misunderstandings, expectations, and decision points. These phenomena are central to interactive news maps (e.g., “I assumed the legend was clickable”, “I did not realise the colours changed after filtering”, “I thought this was an advertisement”). Evidence from eye-tracking studies of election news maps further indicates that verbal reports can be essential for diagnosing representational misunderstandings (e.g., confusion about what choropleth units represent). This can occur even when gaze data show that users attended to the relevant on-screen elements (Popelka et al., 2025). In this sense, combining eye-tracking with talk-aloud supports a more defensible interpretation of gaze patterns and enhances the explanatory power of formative evaluation.

Formative evaluation is particularly suitable for data-journalism maps because news interfaces are often released under tight deadlines. Incremental design improvements can still yield substantial gains in comprehension (Soedirdjo et al., 2025). In cartographic research, formative eye-tracking studies have demonstrated their ability to identify interface elements that users overlook or misunderstand. They also show how these findings can be translated into actionable design recommendations (Çöltekin et al., 2009; Krassanakis & Cybulski, 2021). For example, an eye-tracking evaluation of weather web maps found that users often interact

with them in simple, surface-level ways and may not search for hidden functions. This underlines the importance of discoverability and clear visual cues (Popelka et al., 2019). Similar issues are likely in journalistic environments. Readers are typically not trained to explore the map as an analytical tool, and may prioritise finishing the article over experimenting with interface controls.

This paper reports results from a formative laboratory study initially conducted as a bachelor's thesis (Košťál, 2025). It builds on a formative usability evaluation of interactive maps embedded in data-journalism articles, aiming to identify problematic design and interaction patterns and to formulate practical recommendations for improvement. The study combines eye-tracking with talk-aloud to examine attention allocation and user reasoning when interacting with journalistic maps. The remainder of the paper describes the study design and tasks, summarises the key usability issues observed across cases, and derives practical design recommendations for this class of journalistic map interfaces.

2. Methodology

The study was conceived as a formative usability evaluation of interactive maps embedded in data-journalism articles, combining screen-based eye-tracking with talk-aloud and moderator notes. Data collection took place in the eye-tracking laboratory at the Department of Geoinformatics (Palacký University Olomouc) in November 2024. Eye-tracking evidence was interpreted at two complementary levels: a qualitative review of individual gaze recordings and an AOI-based analysis of page screenshots (e.g., fixation-based attention to text versus graphics), supplemented by scarf plots in GazePlotter (Popelka et al., 2024) to compare viewing patterns across participants. The qualitative inspection of user behaviour was complemented by quantitative usability indicators (e.g., success, efficiency, and error-related patterns).

2.1. Participants

A total of 44 Czech participants were recruited and split into two groups of equal size to enable comparison across levels of domain familiarity: experts (n = 22) and novices/lay readers (n = 22). Experts were defined as participants who work professionally in domains closely related to the communication of socio-demographic data (journalism, sociology, anthropology). Lay participants were defined as typical readers of journalistic articles without specialised expertise in socio-demographic topics. For the sake of anonymity and record traceability, each participant was assigned an internal code (P + ordinal number), which was consistently used in the Tobii Pro Lab recordings and the moderator notes.

2.2. Stimuli and Tasks

Stimuli consisted of five journalistic articles containing interactive map-based (or closely map-related) visualisations, selected after an initial screening and pilot phase. The articles were internally labelled FENTANYL, MIGRATION, UKRAJINA, IROZHLAS, and CNN for analysis and reporting purposes. A preview of the stimuli is clearly shown in Figure 1.

Figure 1. Five journalistic articles for evaluation (three in Czech, two in English)

Each article included 1–2 tasks, operationalised as concrete questions (Q1–Q8 overall), to test whether the interactive element was usable, enabling participants to retrieve the required information through interaction. For two articles (MIGRATION, UKRAJINA), the protocol also included a timed "free exploration" (FE) phase prior to task answering, as shown in Table 1.

Table 1. Overview of tasks for participants (the order of articles was randomised for each participant)

Task ID	Article label	Task	Right answer	Time Limit (s)
Q1	FENTANYL	Which US state has the second-lowest fentanyl mortality rate per 100,000 inhabitants?	Hawaii	120
Q2	FENTANYL	What was the proportion of fentanyl in opioid overdose deaths in 2018?	67%	120
FE1	MIGRATION	Free browsing of the article.	-	240
Q3	MIGRATION	What percentage of the Czech population is immigrants?	5% and 5.1%	120
FE2	UKRAJINA	Free browsing of the article.	-	180
Q4	UKRAJINA	How many inhabitants did the city of Donetsk have before the invasion in 2021?	905,364	120
Q5	IROZHLAS	What was the percentage difference in votes between the most successful and least successful regions in the Karlovy Vary Region for the Five-Party Coalition?	25.3%	120
Q6	IROZHLAS	Which four regions have the highest (worst) destabilising poverty index values in the Karlovy Vary Region?	Aš, Kraslice, Sokolov, Ostrov	120
Q7	CNN	What is the change in percentage points for the Republican Party in Miami-Dade, Florida, in the 2024 presidential election compared to 2020?	18.8 ppt	120
Q8	CNN	Which party won the 2024 presidential election in Los Angeles County (California, on the West Coast) and by how many votes?	The Democrats, with a lead of 1,226,204 votes	120

2.3. Apparatus and Software

Eye movements were recorded using a Tobii Pro Spectrum stationary eye-tracker (sampling rate 300 Hz) controlled in Tobii Pro Lab. Verbal protocols were recorded using an

external microphone (Trust GXT 232, 48 kHz sampling rate). Manual moderator logging was performed on a separate laptop (Lenovo ThinkPad T450S). For post-processing and visual exploration of gaze sequences, the workflow included GazePlotter (v1.6.6) alongside spreadsheet-based analysis.

2.4. Procedure

Firstly, a pilot (n = 3) refined the protocol and timing. In the actual version, participants received standardised instructions and calibration, completed time-limited free exploration (where applicable) and eight information-extraction tasks (120 s each). The talk-aloud approach was also used, including stating participants' answers aloud. A think-aloud protocol was considered but not adopted to keep the procedure lightweight and to preserve a more naturalistic reading and interaction flow. Article order was randomised across participants, and cookie prompts were accepted to preserve realism; moderator intervention was minimal (navigation and technical issues). The complete procedure is shown in Figure 2.

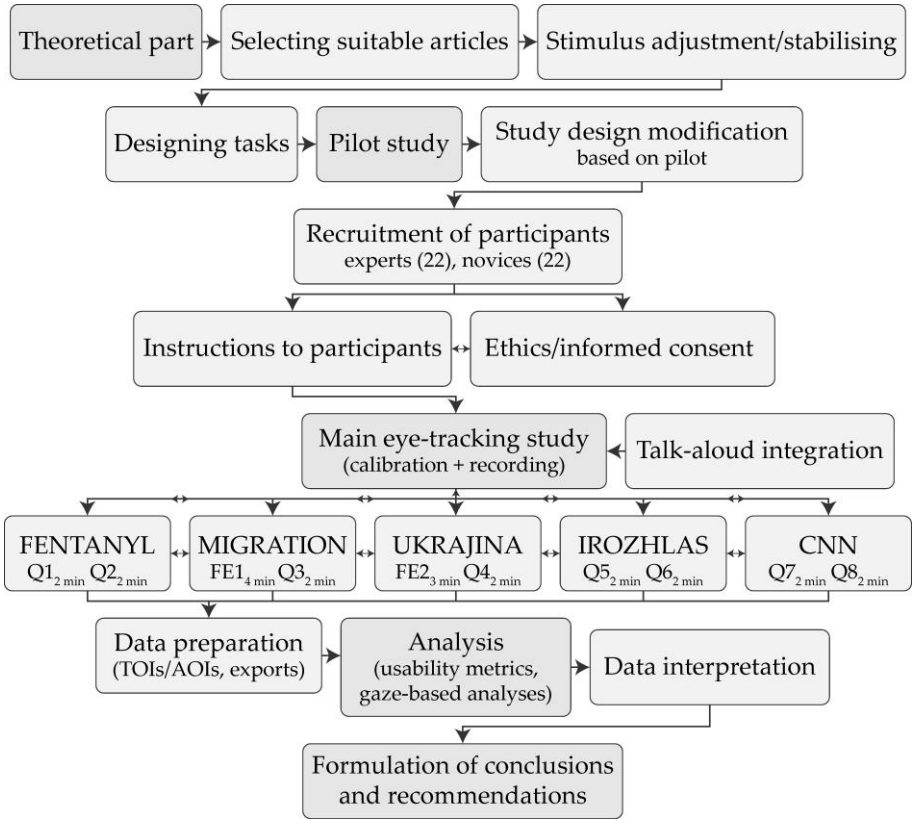

Figure 2. Study the processing procedure

2.5. Measures and Analysis

The analysis combined (1) qualitative behavioural diagnosis (gaze behaviour, interaction strategies inferred from recordings, talk-aloud answers and brief contextual utterances) with (2) quantitative usability indicators. Quantitative indicators were computed at three aggregation levels: task level (Q1–Q8), participant level (performance across all tasks), and group level (experts vs lay users). The expert/lay split was defined a priori (experts:

professionals with a background in socio-demographic data communication; lay users: typical readers without specialised expertise).

Handling of talk-aloud and moderator notes. The talk-aloud approach was used primarily to (i) capture verbal answers for task-success coding and (ii) provide brief contextual cues about the participant's ongoing actions (e.g., intent, uncertainty). The audio was not transcribed, and no formal qualitative codebook, category system, or systematic thematic coding was applied. During each session, the moderator produced brief time-stamped notes (e.g., navigation/technical interruptions and salient behaviours). In post-processing, these notes were used to locate relevant episodes for targeted review alongside screen and gaze recordings in Tobii Pro Lab. Qualitative observations are therefore reported as illustrative, design-relevant examples supporting the behavioural diagnosis rather than as quantified prevalence estimates. The second author (JK) conducted the review of recordings and moderator notes; no independent double-coding or inter-rater reliability assessment was performed.

Task success (effectiveness). For each participant and task, success was coded on a two-point scale: 2 = correct, 1 = incorrect or no answer. Participants who did not provide a verbalised answer within the time limit were treated as having no answer (= 1). Task-level success rates were computed as the percentage of correct responses per task; participant-level success was calculated as the percentage of correctly solved tasks per participant.

Task completion time (efficiency). Each information-extraction task had a fixed time limit of 120 s (Table 1). Time-on-task was measured in Tobii Pro Lab using task-specific TOIs defined from task presentation to the verbalised answer. If a participant did not answer (or did not manage to verbalise an answer) within 120 s, the time-on-task was set to 120 s (i.e., a fixed cap imposed by the protocol time limit).

Empirical task difficulty (D). Because talk-aloud was mainly used as a supplementary interpretive aid and many participants verbalised only minimally, verbal statements were not treated as a reliable basis for estimating task difficulty. Difficulty was therefore derived empirically from mean task success (lower success indicates greater difficulty) and discretised into terciles (D: 1 = most difficult, 3 = most manageable).

Composite respondent score (efficiency beyond correctness). To summarise performance beyond correctness alone, we computed a composite respondent score combining success, empirical difficulty, and time-on-task weighting across Q1–Q8:

$$Score_i = \sum_{j=1}^{8} \left(\frac{S_{ij}}{D_j} \times T_{ij} \right) \quad (1)$$

where S_{ij} is task success (2/1), D_j is task difficulty tercile (1–3), and T_{ij} is a within-task speed tercile for participant i on task j (3 = fastest, 1 = slowest), computed separately for each task from the distribution of completion times. This weighting rests on two protocol-driven assumptions: (a) task difficulty can be approximated by empirical success within the present sample, and (b) a uniform time limit (120 s) makes time-on-task comparable across

tasks under a capped measurement scale. The score, therefore, rewards correct answers, assigns greater credit to correctness on empirically more complex tasks (via $1/D_j$), and differentiates efficiency within each task (via terciles), reducing sensitivity to skew and outliers. The maximum attainable score under this scheme is 29 points.

Error-related indicator (repeated-question corrections). As a diagnostic complement, we tracked cases in which a question was repeated after an initial incorrect response, allowing participants to correct their answer. This indicator served to distinguish potential misunderstanding of the question/context from visualisation-driven failure and was summarised descriptively at task, participant, and group levels.

Group comparisons (experts vs lay users) were treated as descriptive and exploratory (e.g., comparing distributions/variability alongside means), because the primary goal was to derive design-focused insights rather than to make strong inferential claims about user-group differences.

Eye-tracking analysis was combined with a qualitative review of recordings, utilising AOI-based metrics computed on page screenshots (including higher-level AOI categories for free exploration). Exported visited AOI sequences were inspected in GazePlotter (Popelka et al., 2024) to compare attention-switching patterns and to guide targeted review of cases with incorrect or missing answers. AOIs were delineated on Tobii Pro Lab–generated page screenshots using an article-specific AOI scheme. They were drawn to cover whole functional elements (e.g., text blocks, graphs, and maps) irrespective of their on-screen size. Fixations were identified in Tobii Pro Lab (v1.241.54542) using the I-VT Gaze Filter with default settings, and AOI metrics were exported as TSV for scarf-plot inspection in GazePlotter.

3. Results

Across all eight tasks (Q1–Q8), the mean task success was 57.67%. Task-level success rates varied substantially, with Q3 (79.55%) and Q6 (77.27%) among the most successful tasks, and Q8 (20.45%) and Q1 (34.09%) among the least successful. At the participant level, P33 (lay user) solved all tasks correctly (100%). The least successful participants were P18 (expert) and P22 (lay user), with the same success rate (12.50%).

Because talk-aloud was used mainly as a supplementary interpretive aid and many participants verbalised only minimally, the study did not treat talk-aloud statements as a reliable basis for estimating task difficulty. Difficulty was instead derived from mean task success and then classified into terciles. Based on the tercile classification, Q1, Q5, and Q8 were identified as the most challenging tasks, whereas Q3, Q4, Q6, and Q7 were the most manageable tasks.

Each task had a fixed time limit of 120 seconds, and time-on-task was measured using TOIs in Tobii Pro Lab (from task presentation to verbalised answer). For participants who did not provide (or did not manage to verbalise) an answer within the 120-second limit, time-on-task was recorded as 120 s, and the task was coded as unsuccessful (success = 1). Mean completion times (seconds) varied notably across tasks, with the overall pattern consistent with the success results (longest: Q8 = 110.99 s, also the least successful task; shortest: Q6 = 56.83 s, among the most successful tasks). This co-occurrence (low success

+ long time) may reflect higher cognitive demands and/or lower comprehensibility of the interactive solution for that task.

To complement success rates, we computed a composite respondent score combining task success, task difficulty, and time-on-task weighting across Q1–Q8. The top performers were P03 (expert) and P33 (lay) (24.33 points each). Notably, P03 ranked only fourth by raw success (7/8 correct) but achieved a top score due to faster performance on more demanding tasks, indicating that efficiency meaningfully differentiated participants beyond correctness.

At the group level, the mean difference between experts and lay users was negligible; however, the distributions differed: experts showed lower variance (greater concentration around the mean), while lay users exhibited higher variability in performance.

We additionally tracked repeated-question corrections to distinguish misunderstandings of the question from visualisation-driven failures. This indicator is compatible with some initial errors being contextual (e.g., article structure or time pressure), while low-performing tasks were frequently associated with weakly discoverable interaction.

For the free-exploration segment, we quantified attention to text versus graphics using AOI-based metrics. After normalising by pixel area, participants still allocated markedly closer relative attention to text than to graphics (lay users: 10.91×; experts: 17.46×), indicating intense competition between narrative content and visual elements in news-page layouts.

Scarf plots in GazePlotter were then used to compare attention-switching patterns among participants (Figure 3) and to identify instances in which users did not reach the intended interactive element within the time limit, which was verified through a targeted review of the corresponding recordings.

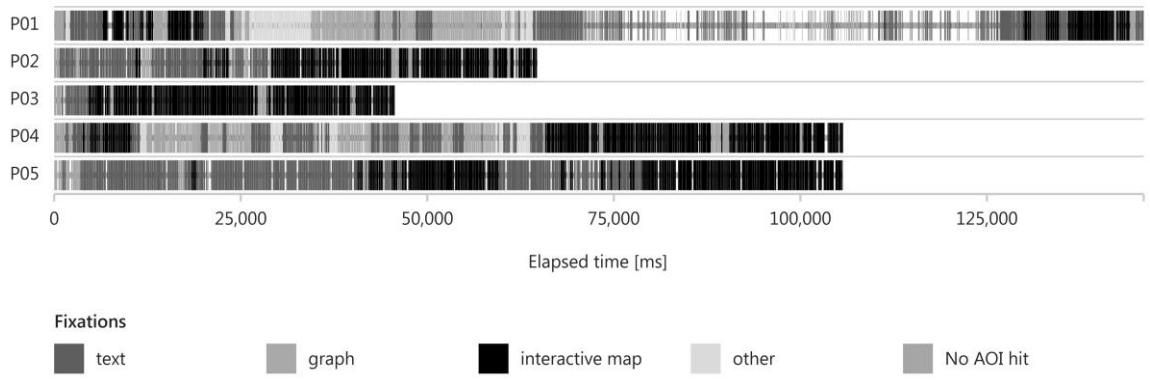

Figure 3. Example of using a scarf plot for Q3 in GazePlotter to compare data from the first five participants (the black-and-white version shown here is provided for illustrative purposes only)

Beyond aggregated metrics, the study highlights several recurring usability issues derived from observed behaviour:

- Interactivity not recognised/insufficient interaction with controls. Some participants did not use the mouse enough, or did not recognise an element as interactive at all.
- Secondary/hidden interaction mechanisms produced major failures (notably Q8). For the problematic Q8 task, most participants did not answer correctly on the first attempt and tended to use only the default map layer; the switching control and the secondary

map accessible via it showed weak interactive potential and limited usability for extracting the required value.

- Language and time constraints. In the same scenario, part of the difficulty was attributed to possible misunderstandings of the English text; some participants appeared to locate the correct answer but did not verbalise it before the time limit, leading to no response.

These findings align with the overall quantitative pattern: tasks with poorly discoverable or cognitively demanding interactive mechanisms yielded the lowest success rates and the longest completion times, whereas tasks with clearer interaction affordances (or where the answer could be reliably derived without complex interaction) achieved higher success rates with shorter completion times. Interactive news maps can support targeted information extraction; however, in this study, usability degraded quickly when interactivity was not discoverable, when interactive components were placed low in the narrative, or when interaction feedback was fleeting.

4. Discussion

The results suggest that usability depended primarily on the legibility and placement of interaction mechanisms, not on the mere presence of a map. Failures concentrated in tasks requiring secondary or weakly signposted controls (e.g., layer switching), suggesting that the key barrier was interaction discoverability rather than map reading. Further free exploration reveals intense competition between narrative text and graphics, suggesting that interactive maps may not be used, especially when placed low in the article. When static and interactive maps coexisted, users were often satisfied with the first acceptable (static) representation, underscoring the need to signal why interaction is necessary and to position interactive elements early.

The findings translate into three design implications for data-journalism maps: (1) Make interactivity explicit (e.g., visual cues for clickable/switching controls) so that key mechanisms are noticed in seconds. (2) Place interactive elements early and avoid competing static alternatives or clearly explain why interaction adds value. (3) Support micro-interactions by reducing transient feedback (e.g., longer-lasting pop-ups) and by providing navigational aids (anchors) in longer stories so readers can reach the relevant interactive element efficiently.

At the group level, mean performance was similar for experts and lay users; given the primarily student sample and a potentially weak operationalisation of expertise, the data mainly support design-focused conclusions rather than claims about user-group differences.

The composite respondent score should be interpreted as a relative indicator tailored to this protocol: difficulty is estimated from empirical success in the current sample (terciles), and the 120-second cap bounds completion time. Alternative weightings (e.g., continuous time, external difficulty estimates, or different timeout penalties) may change absolute scores. Still, the metric was introduced to provide a more discriminative view of efficiency than correctness alone for this set of tasks.

Interpretation should consider four constraints: (1) gaze indicates attention, not understood/used interactivity so that interaction logs would improve diagnosis; (2) interactive map states vary across users, limiting cross-participant comparability of gaze metrics; (3) news-web stimuli introduce procedural noise (links, cookies, loading/rendering), motivating stronger stimulus stabilisation (e.g., local hosting and link control), and (4) talk-aloud and moderator notes were collected mainly to capture verbal answers and minimal context, and were not analysed through a formal coding procedure; consequently, the qualitative insights should be interpreted as illustrative.

Future work should (a) triangulate gaze with interaction logs to separate "seen" from "used" controls and (b) stabilise web stimuli (e.g., local hosting/link control), while (c) recruiting more clearly defined expert samples to test group effects with inferential statistics.

5. Conclusions

This study evaluated the usability of interactive news maps embedded in data-journalism articles using laboratory eye-tracking complemented by talk-aloud and moderator observations. Across eight information-extraction tasks, overall effectiveness was moderate (mean success 57.67%) and varied strongly by task. Efficiency measures (time-on-task) aligned with success patterns, indicating that failures were structured and linked to specific interaction demands.

The results point to a central usability bottleneck: the discoverability of interactions and feedback. When tasks required secondary or weakly signposted controls, many participants did not detect the affordance or could not complete the necessary interaction within the time limit. Free-exploration analyses further show that readers devoted substantially more attention to text than to graphics, even after normalising by pixel area, suggesting that interactive maps compete with narrative content and page clutter and cannot be assumed to be used simply because they are present.

Overall, for interactive news maps, usability is primarily a problem of interaction design and narrative integration—not merely a cartographic one.

Conflict of interest: none

Acknowledgements: This work was supported by GAČR – Czech Science Foundation under the "Identification of barriers in the process of communication of spatial socio-demographic information" (23-06187S) project. The work of Tomas Vanicek was supported by the Internal Grant Agency of Palacký University Olomouc (grant number IGA_PrF_2025_002).

References

Blascheck, T., Vermeulen, L.M., Vermeulen, J., Perin, C., Willett, W., Ertl, T., & Carpendale, S. (2019). Exploration strategies for discovery of interactivity in visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *25*(2), 1407–1420. https://doi.org/10.1109/TVCG.2018.2802520

Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S.I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17. https://doi.org/10.1559/152304009787340197

Ericsson, K.A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P.J. Feltovich, & R.R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223–242). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.013

Fairbairn, D., & Hepburn, J. (2023). Eye-tracking in map use, map user and map usability research: What are we looking for? *International Journal of Cartography*, *9*(2), 231–254. https://doi.org/10.1080/23729333.2023.2189064

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

International Organization for Standardization. (2018). *ISO 9241-11: Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*. ISO.

Karypidou, C., & Veglis, A. (2019). Visualization and interactivity in data journalism projects. *Strategy and Development Review, 9*(Special Issue), 44–60. https://doi.org/10.34276/1822-009-999-004

Košťál, J. (2025). *Formativní analýza interaktivních map v datové žurnalistice* (Bakalářská práce, Univerzita Palackého v Olomouci). https://www.geoinformatics.upol.cz/dprace/bakalarske/kostal25/

Krassanakis, V., & Cybulski, P. (2021). Eye tracking research in cartography: Looking into the future. *ISPRS International Journal of Geo-Information*, *10*(6), 411. https://doi.org/10.3390/ijgi10060411

Molina Rodríguez-Navas, P., Muñoz Lalinde, J., & Medranda Morales, N. (2021). Interactive maps for the production of knowledge and the promotion of participation from the perspective of communication, journalism, and digital humanities. *ISPRS International Journal of Geo-Information*, *10*(11), 722. https://doi.org/10.3390/ijgi10110722

Popelka, S. (2018). *Eye-tracking (nejen) v kognitivní kartografii: Praktický průvodce tvorbou a vyhodnocením experimentu*. Univerzita Palackého v Olomouci.

Popelka, S., Dorusakova, J., Vanicek, T., & Vojtechovska, M. (2025). Mapping democracy: An eye-tracking analysis of election maps. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, Article 82. https://doi.org/10.1145/3715669.3725865

Popelka, S., Kominek, J., & Vojtechovska, M. (2024). Exploring geological map usability through sequence chart visualization. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications (ETRA '24)*. Association for Computing Machinery, New York, NY, USA, Article 79. https://doi.org/10.1145/3649902.3653520

Popelka, S., Vondrakova, A., & Hujnakova, P. (2019). Eye-tracking evaluation of weather web maps. *ISPRS International Journal of Geo-Information*, *8*(6), 256. https://doi.org/10.3390/ijgi8060256

Rafeeq, A. (2025). Multimedia storytelling in online journalism: Analysing multimedia use on the websites TheGuardian.com, TheNationalnews.com, USAToday.com, and Stuff.co.nz. *Journalism and Media*, *6*(3), 157. https://doi.org/10.3390/journalmedia6030157

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Soedirdjo, R.B., Kusumo, A.H., & Hartono, M. (2025). Improvements to visual design on the online news portal ABC.com by employing usability and eye tracking methods. *Engineering Proceedings*, *84*(1), 72. https://doi.org/10.3390/engproc2025084072

Stalph, F. (2018). Classifying data journalism. *Journalism Practice*, *12*(10), 1332–1350. https://doi.org/10.1080/17512786.2017.1386583

Vanicek, T., Beitlova, M., Vojtechovska, M., & Popelka, S. (2024). Exploring methods for evaluation of dynamic and interactive maps. In *Proceedings of the 9th International Conference on Cartography and GIS* (pp. 1–9). https://iccgis2024.cartography-gis.com/papers/9ICCGIS-Proceedings_Paper%20%2827%29.pdf

Vanicek, T., & Popelka, S. (2023). The think-aloud method for evaluating the usability of a regional atlas. *ISPRS International Journal of Geo-Information*, *12*(3), 95. https://doi.org/10.3390/ijgi12030095

Zbiejczuk Suchá, L., Popelka, S., Šašinková, A., Godišková, G., Beitlová, M., Vaníček, T., & Šašinka, Č. (2026). The map maker's dilemma: Design challenges in interactive mapping for social science and data journalism. *Information Discovery and Delivery*, 1–14. https://doi.org/10.1108/idd-11-2024-0184

Zhang, B., Zhang, Y., & Pan, Y. (2023). Accepting the digital challenge: Public perceptions and attitudes toward interactivity in data journalism. *Applied Sciences*, *13*(19), 10857. https://doi.org/10.3390/app131910857