

Video Replay of Eye Tracking as a Cue in Retrospective Protocol

...Don't Make Me Think Aloud!

by Lizzie Maughan, Jon Dodd and Richard Walters

Summary

We evaluate the effectiveness of Post Experience Eye-tracking Protocol (PEEP) as a web usability testing technique. Popular think aloud protocols often interfere with cognitive processes during task completion. In contrast, retrospective verbalization allows for natural behaviour but is subject to fallibilities associated with memory retrieval and the possible editing or fabrication of thoughts which can occur post experience. To overcome these shortcomings we show a video replay of gaze patterns back to the user to cue verbalization. We find that a significantly greater percentage of users complete their set task when using PEEP. PEEP also leads to a greater reporting of usability problems. Moreover, we find that it is the replay of eye gaze patterns *per se* that cues a greater reporting of usability problems rather than a review of solely screen images and cursor movements.

Current usability testing protocols

Think Aloud

Think Aloud protocols are widely used in usability research. The technique requires the user to verbalize their actions and thoughts as they navigate a website (Ericsson & Simon 1993; Nielsen & Yssing 2004). Though it has high face validity, limitations include:

- Users frequently find the verbalisation unnatural, difficult and unpleasant.
- Users behaviour is influenced due to introspection.
- Users may not report cognitive processes that are difficult to translate verbally or which occur unconsciously.

Retrospective verbalization

Though less popular than Think Aloud, retrospective protocols are regarded as equally effective (Neilson 1993; van den Haak et al. 2003). Retrospective verbalization can overcome many of the problems associated with Think Aloud but has two key limitations:

- Reliance on long term memory: human encoding, storage and retrieval processes are fallible.
- Thoughts maybe subject to post-hoc rationalizations, biases and fabrication.

An alternative method

Post Experience Eye-tracking Protocol (PEEP)

PEEP is similar in principle to eye mark retrospections (Hansen 1990). Video replay of tracked eye movements act as a cue to retrieve thoughts that occurred during the completion of a prior task. PEEP utilizes modern eye tracking hardware and software to refine this technique and overcome many previously identified limitations (Cooke & Cuddihy 2005; Bednarik & Tukiainen 2006). Critically, an experienced moderator can recognize, in real time, distinct gaze patterns correlated to particular behaviours which can then allow for a more in-depth probing of the user's underlying motivations post experience.

Key stages of PEEP

Key stages of PEEP	
1	User is calibrated for eye tracking.
2	User can be asked to complete a task on their own, without intervention and without being asked to voice their thoughts.
3	Moderator watches user eye traces in real time and notes any interesting gaze patterns. Monitoring can be conducted from a separate room.
4	On completion of the task the moderator asks the user to comment on the dynamic replay of their eye-movements. The user is shown both video highlights and static representations of their eye traces.
5	Moderator can show any distinct eye gaze patterns to probe further the underlying motivation of the user.

The advantages of PEEP over Think Aloud

Because behaviour is more natural eye-tracking can reveal underlying mental processes that would otherwise be obscured by Think Aloud (see fig. 2).

- Eye trace of user can be observed in real time.
- Video replay of eye-track data acts as an objective cue for moderators to collect qualitative information from users post experience.
- Retrospective verbalization provides a richer, more cohesive discussion for analysis.
- More realistic task complete rate.
- Users report a greater number of usability problems.
- Distinct gaze signatures can be identified by an experienced moderator and used to further investigate underlying user motivation.

Bunnyfoot has identified a dozen such gaze signatures correlated to behaviours such as reading, decision making and search.

Validation of PEEP

Experimental design

Nicola Egar and Linden Ball at the Department of Psychology at Lancaster University conducted a nested experimental design with 24 participants (mean age = 26 years; 1:1 sex ratio). Participants were asked to view two alternative websites *Google* and *Infomagnet* under *Think Aloud* versus *PEEP* monitoring conditions. Experimentalists controlled for input words and search results and within and between participants effects. Two different cues were used for retrospective verbalization: viewing of the screen (with cursor movements) vs. viewing of the screen with recorded eye traces from the eye tracker superimposed. Questionnaires were completed both prior to testing (demographic data) and post testing (views on their overall experience, working method and the presence of the experimenter; based on a five point Likert scale) (Ball et al. 2006).

Key results

- A significantly greater total number of usability problems were identified by users under PEEP (fig. 1a). Specifically, PEEP led to a greater reporting rate of problems relating to comprehension of search results and feedback from the website. There was no significant difference in the reporting of problems related to layout, terminology or navigation.
- The use of screen images solely as a cue did not lead to significant increase in usability problems being reported over Think Aloud. Thus eye traces *per se* were found to be important in identifying usability problems.
- A significantly greater proportion of users successfully completed their task under PEEP than Think Aloud (fig. 1b).
- Users felt that their working conditions under PEEP were both faster ($P=0.015$) and more focussed ($P<0.001$) than under Think Aloud.
- Users felt that the presence of the experimenter was significantly less unpleasant ($P=0.006$) and unnatural ($P=0.031$) under PEEP than Think Aloud.

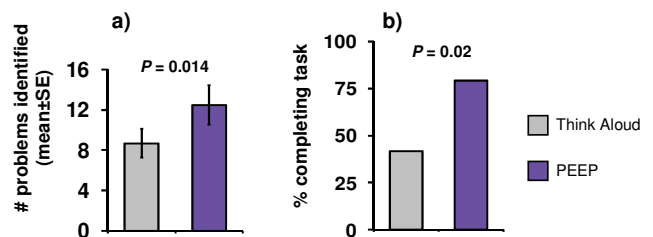


Figure 1. Comparison of usability user techniques in regard to a) eliciting usability problems from users and b) success rate of users completing their set task.

...which is more natural?

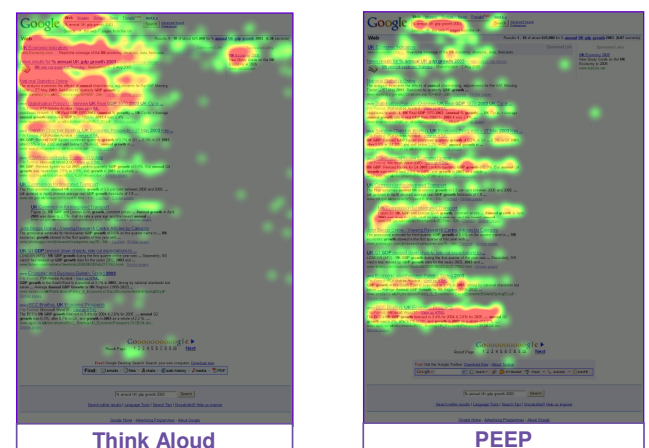


Figure 2. 'Heat map' visualizations showing the mean sum of fixation points among users recorded during completion of a task under alternative usability user testing techniques: Thinking Aloud (Pretorius, Calitz & Greunen 2005) versus PEEP.

References

Bednarik & Tukiainen (2006) *An eye-tracking methodology for characterizing program comprehension processes*. ETRA, San Diego, California. Cook & Cuddihy (2005) *Using Eye Tracking to Address Limitations in Think-Aloud Protocol*. IEEE International Professional Communication Conference Proceedings. Ball, Eger, Stevens & Dodd (2006) *Applying the Post-Experience Eye-Tracked Protocol (PEEP) method in usability testing*. *Interfaces*, 67, 15-19. Ericsson and Simon (1993) *Protocol Analysis: Verbal Reports as data*. Cambridge, MA: MIT Press. Hansen (1990) *The use of eye mark recordings to support verbal retrospection in software testing*. *Acta Psychologica* 76: 31-49. Nielsen (1993) *Usability Engineering*. Boston MA, Academic Press. Nielsen & Yssing (2004) *What Kind of Information does an HCI expert want?* NORDICHI, Tampere, Finland. Pretorius, Calitz & Van Greunen (2005) *The Added Value of Eye Tracking in the Usability Evaluation of a Network Management Tool*. SAICSIT 1-10. van den Haak et al. (2003) *Retrospective vs. concurrent think aloud protocols: testing the usability of an online catalogue*. *Behaviour of Information Technology* 22: 339-351.