

Comparing of Feedback Collection and Think-Aloud Methods in Program Comprehension Studies

Amela Karahasanovic, Unni Nyhamar Hinkel, Dag I. K. Sjøberg and Thomas Richard

This paper reports an explorative experimental comparison of (i) an experience-sampling method called feedback collection and (ii) the think-aloud methods with respect to their usefulness in studies on program comprehension. Think-aloud methods are widely used in studies of cognitive processes, including program comprehension. Alternatively, as in the feedback-collection method (FCM), cognitive processes can be traced by collecting written feedback from the subjects at regular intervals. We compare FCM with concurrent think-aloud (CTA) and retrospective think-aloud (RTA) regarding type and usefulness of the collected information, costs related to analysis of the collected information, and effects of the data collection methods on the subjects' performance. FCM allowed us to identify a greater number of comprehension problems that prevented progress or caused significant delay (FCM: 30 problems; CTA: 5; RTA: 15). It was less precise in identifying strategies for comprehension than CTA (92% correctness for FCM; 100% for CTA). FCM was less expensive in analysis (transcription and coding) than the other two methods (FCM: 0,7 hours of analysis per protocol; CTA: 31 hours; RTA: 7,9 hours). The results indicate that all three methods of data collection were intrusive and affected the performance of the subjects with respect to time and correctness (small to medium effect size). This research confirms that FCM can be used beneficially in studies that trace the cognitive processes involved in, and identify problems related to, the comprehension of software applications. Based on our experience, we recommend that FCM be used in studies that have a large number of subjects and as a complement to other methods for tracing cognitive processes, such as user log files. We recommend a design with two groups (verbalization and silent control) and a pretest task to be used in studies with FCM or CTA that focus on performances.