Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting

DOI:
10.1037/1076-898X.11.4.237

Document status and date:
Published: 01/12/2005

Journal of Experimental Psychology: Applied
2005, Vol. 11, No. 4, 237–244

1076-898X/05/$12.00 DOI: 10.1037/1076-898X.11.4.237

Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting

Tamara van Gog, Fred Paas, Jeroen J. G. van Merriënboer, and Puk Witte
Open University of the Netherlands

This study investigated the amounts of problem-solving process information ("action," "why," "how," and "metacognitive") elicited by means of concurrent, retrospective, and cued retrospective reporting. In a within-participants design, 26 participants completed electrical circuit troubleshooting tasks under different reporting conditions. The method of cued retrospective reporting used the original computer-based task and a superimposed record of the participant's eye fixations and mouse–keyboard operations as a cue for retrospection. Cued retrospective reporting (with the exception of why information) and concurrent reporting (with the exception of metacognitive information) resulted in a higher number of codes on the different types of information than did retrospective reporting.

Keywords: process-tracing techniques, verbal reports, eye tracking, problem solving, knowledge elicitation

Process-tracing methods, such as concurrent reporting ("thinking aloud"), retrospective reporting, eye tracking, and decision analysis, can be used to elicit information that allows for making inferences about the cognitive processes underlying problem-solving performance (Cooke, 1994). Hence, these methods are widely applied, for example, in usability studies, to investigate how people interact with a system or device in order to improve it (e.g., van den Haak, De Jong, & Schellens, 2003); in the design of expert systems, to uncover expert cognitive processes in order to model a system (see Richman, Gobet, Staszewski, & Simon, 1996); and in educational research, either to uncover problem-solving processes as a goal in itself or to improve instruction (e.g., Renkl, 1997).

Problem solving is defined as getting from an initial problem state to a desired goal state, without knowing exactly what actions are required to get there (Newell & Simon, 1972). In problem solving, different types of knowledge are applied. *Domain knowledge* (principles) is used to mentally represent the problem and to narrow down the problem space to those problem-solving operators (i.e., solution steps or actions) that may be relevant for this kind of problem. It interacts with *strategic knowledge* (heuristics, systematic approaches to problem solving), which is used to select operators that are most likely to lead to the goal state. *Metacognitive knowledge* is used to monitor this process of selection and application of operators by keeping track of the progress toward the goal state.

For the design of instruction that makes all the knowledge used in the problem-solving process explicit to learners (e.g., process-oriented worked examples; see van Gog, Paas, & van Merriënboer, 2004), a process-tracing technique is required that is able to uncover information about problem-solving actions taken ("action"), domain principles used ("why"), strategies used ("how"), and self-monitoring ("metacognitive"). However, the results obtained with the two most widely applied verbal methods, concurrent reporting and retrospective reporting, suggest that neither of these methods is suitable for providing a comprehensive picture of the problem-solving process in terms of action, why, how, and metacognitive information.

With the method of concurrent reporting (Ericsson & Simon, 1993; van Someren, Barnard, & Sandberg, 1994), participants are instructed to think aloud, that is, to verbalize all thoughts that come to mind while working on a task (i.e., online). In retrospective reporting (Ericsson & Simon, 1993), participants are instructed to report the thoughts they had while they were working on a task immediately after completing it (i.e., off-line). It should be noted that in order to allow for valid inferences about the cognitive processes underlying task performance, the wording of verbalization instructions and prompts is crucial. Only when the instructions and prompts are worded in such a way that the evoked responses do not interfere with the cognitive processes can concurrent and retrospective reporting result in verbal protocols that reflect the reported cognitive processes (Ericsson & Simon, 1993).

However, there is an important distinction between the information contained in concurrent and retrospective protocols, related to their online and off-line generation, respectively. Concurrent protocols reflect the information available in short-term memory during the process, whereas retrospective protocols reflect the memory traces of the process that are retrieved from short-term memory (in tasks of very short duration) or long-term memory directly after it is finished (Camps, 2003; Ericsson & Simon,

Tamara van Gog, Fred Paas, Jeroen J. G. van Merriënboer, and Puk Witte, Educational Technology Expertise Center, Open University of The Netherlands, Heerlen, the Netherlands.

This research project is funded by the Netherlands Organization for Scientific Research (NWO, The Hague, Project No. 411-01-010). We gratefully acknowledge K. Lynn Taylor's efforts at providing the effect sizes for the data reported in Taylor and Dionne (2000).

Correspondence concerning this article should be addressed to Tamara van Gog, Educational Technology Expertise Center, Open University of the Netherlands, P.O. Box 2960, 6401 DL, Heerlen, the Netherlands. E-mail: tamara.vangog@ou.nl

1993). This reference to different memory systems seems to result in differences in the problem-solving information contained in the protocols. For example, Taylor and Dionne (2000) noted that concurrent protocols seem to predominantly contain information on actions and their outcomes, whereas retrospective protocols seem to contain more "references to strategies that control the problem-solving process" and "information such as the conditions that elicited a particular response" (cf. our categories of why, how, and metacognitive information; Taylor & Dionne, 2000, p. 414). They reported only means and standard deviations, but an analysis of their data (all participants, $N = 36$) shows that the number of codes on their category of actions was significantly higher in concurrent protocols, with a large correlation effect size, *r* (Rosenthal, Rosnow, & Rubin, 2000), of .84, and that the number of codes on their categories of "conditional knowledge," "beliefs," and "strategy acquisition knowledge" (cf. our categories of why, how, and metacognitive information) was significantly higher in retrospective protocols (large correlation effect sizes, *r,* of .91, .93, and .94, respectively). Kuusela and Paul (2000) found that concurrent protocols contained more information than retrospective protocols, because the latter often contained only references to the effective actions that led to the solution.[1] This might be a result of participants' selective reporting of the correct solution steps to represent their performance as better than it actually was, but a more likely explanation is that only the correct steps, that have led to attainment of the goal, are stored in long-term memory, as only these steps are relevant for future use. So, because retrospective reporting requires retrieval of episodic memories from long-term memory, reports can be subject to forgetting, which may explain why less information on actions is elicited with this method. Besides forgetting, another problem with retrospective reports is fabricating, that is, off-line reporting of information that was not actually part of the online process. It is unclear whether this might explain the finding that retrospective protocols seem to contain more why, how, and metacognitive information, as this knowledge might also have been used during the process but might be omitted in concurrent reporting as a result of the greater processing demands of this method (see Russo, Johnson, & Stephens, 1989). What is clear, however, is that neither of these methods seems able to provide a comprehensive picture of the problem-solving process in terms of action, why, how, and metacognitive information.

A technique is needed that combines the advantages of concurrent reporting (i.e., more action information) and retrospective reporting (i.e., more why, how, and metacognitive information). Of course, one could use both methods complementarily (Camps, 2003; Ericsson & Simon, 1993; Taylor & Dionne, 2000). However, there is a methodological risk involved in retrospective reporting on the same task that was carried out while reporting concurrently, because the episodic memory retrieved and reported might be the memory of the concurrent verbalizations instead of the memory of the process. We propose that a method of cued retrospective reporting, by using a combined record of eye movements and mouse–keyboard operations that is superimposed on the problem as a cue, might be able to combine the advantages of both other methods.

In cued retrospective reporting, participants are instructed to report retrospectively on the basis of a record of observations or intermediate products of their problem-solving process, which they are shown to cue their memories of this process. This is known to lead to better results because of less forgetting and/or fabricating of thoughts than plain retrospective reporting (van Someren, et al., 1994). More important for our purposes, cued retrospective reporting based on a cue that shows participants' actions might lead to more actions being reported, without losing the retrospective nature and its associated information types. A videotape of the problem-solving session can be used as a cue of actions, but for computer-based tasks in domains in which visual inspection plays a key role (e.g., troubleshooting electrical circuit simulations, numerically controlled machinery programming), a cue consisting of a combined record of eye movements and mouse–keyboard operations that is superimposed on the problem can be expected to lead to better results. Because eye movements reflect cognitive processes (Lauwereyns & d'Ydewalle, 1996; Rayner, 1998), they might cue the participants to report on those processes, including visual problem-solving actions that would not be observable on videotape or in a record of mouse–keyboard operations. For example, when troubleshooting electrical circuit simulations, the problem-solving process would (ideally) start with surveying the circuit and diagnosing possible malfunctioning components. Such activities would become visible only in a record of eye movements, and using such a record superimposed on the problem-solving task as a cue can lead to these actions and the underlying why and how knowledge being reported, even if this concerns implicit (tacit) knowledge (Lauwereyns & d'Ydewalle, 1996).

Support for the assumptions with regard to this type of cue comes from the work of Russo (1979) and Hansen (1991). Although Russo (1979) did not distinguish between different types of information, he found that retrospective reporting based on a record of eye fixations resulted in protocols that contained more words and were of longer duration than concurrent and retrospective protocols.[2] Russo et al. (1989) conducted a study that might have provided interesting information on the use of eye movements as a cue. They intended to compare errors of fabricating and forgetting in different types of retrospective reporting: on the basis of the initial stimulus, the response, or a record of eye movements. Unfortunately, however, the use of an improper instruction for retrospective reporting based on eye movements made comparisons between this type and the other two types pointless. Hansen (1991) compared the number of cognitive, manipulative (physical actions on the PC, e.g., typing), and visual operational (visual actions on the PC, e.g., reading) comments elicited with retrospective reporting cued by either a record of eye movements or a video recording. Although both methods resulted in the same number of comments on cognitive operations, the condition with the video record as a cue resulted in the highest number of manipulative comments, whereas the condition with the eye movement record as a cue yielded more visual operational comments. So, for computer-based tasks that rely on visual inspection, a combined record of

[1] Kuusela and Paul (2000) used a between-participants design; however, because they did not report standard deviations, effect sizes could not be computed on the basis of the data presented in that article. Our attempts to obtain the required data were unsuccessful.

[2] Russo (1979) used a within-participants design, and hence, effect sizes could not be computed on the basis of means and standard deviations presented in that article. Our attempts to obtain the required data were unsuccessful.

mouse–keyboard operations and eye movements is expected to lead to the best results, because it can trigger memory of physical as well as visual cognitive actions.

The question addressed in this study concerns the differences in the amounts of action, why, how, and metacognitive information elicited with the methods of concurrent reporting, retrospective reporting, and cued retrospective reporting. The tasks used are computer-simulated electrical circuits troubleshooting problems. It is expected that concurrent protocols would contain more action information than retrospective protocols (cf. Kuusela & Paul, 2000; Taylor & Dionne, 2000) and that retrospective protocols would contain more why, how, and metacognitive information than concurrent protocols (cf. Taylor & Dionne, 2000). By combining the advantages of concurrent and retrospective protocols, cued retrospective protocols are expected to reveal the most comprehensive picture of the problem-solving process by (a) providing more action information than retrospective protocols and (b) providing more why, how, and metacognitive information than concurrent protocols.

Method

Design

A within-participants design with four conditions was used: concurrent reporting, retrospective reporting, cued retrospective reporting, and concurrent reporting with eye tracking. Each condition was paired with two tasks, and the conditions were counterbalanced, resulting in four sequences to which participants were randomly assigned. The concurrent reporting with eye-tracking condition was added to gather data outside the scope of this article. Note, therefore, that other data from the participants in this study are reported in van Gog, Paas, and van Merriënboer (2005).

Participants

Participants were 26 students (17 male, 9 female; age range = 17–21 years) in their 5th year of pre-university education or their 1st or 2nd year of higher professional education, all of whom had uncorrected good eyesight or good eyesight when corrected with hydrophilic contact lenses. They had studied at least the basic theory of electricity, so each participant knew how circuits and the individual components should function. They volunteered to participate and received €12.50 after the experiment. (Note: This paragraph describes information also present in van Gog et al., 2005, p. 209.)

Apparatus and Materials

Registration of eye movements. A 50-Hz video-based remote eye-tracking device from SensoMotoric Instruments (iView manual version 3.01, IV-DOC-30-X; Teltow, Germany) was used to record eye movements. This camera with infrared source was placed under the 21-in. (53.34-cm) screen of the stimulus PC, located in the recording room. By means of an adjustable forehead rest that was placed in front of the screen, participants were positioned so that their eyes were approximately 70 cm from the screen's center. On a connected PC in the adjoining observation room, iView software (SensoMotoric Instruments) was used to operate the camera and to calibrate the eye-tracking system. To enable the experimenter to perform the necessary actions on the stimulus PC when calibrating the system from the observation room, an extra mouse, keyboard, and monitor located in the observation room were connected to the stimulus PC. Participants' eye movements and mouse and keyboard operations were registered and replayed by using GazeTracker software (Lankford, 2000). A one-way screen enabled the experimenter to observe the recording room from the observation room, and microphones enabled verbal communication between both rooms. These microphones were attached to a digital audiorecorder to enable recording of participants' verbal reports. (Note: This paragraph describes information also present in van Gog et al., 2005, p. 212.)

Troubleshooting tasks. The troubleshooting tasks consisted of computer-simulated malfunctioning electrical circuits. They were constructed by a science teacher in the Crocodile Physics 1.5 software program and were at a level of difficulty appropriate for 4th-year higher general secondary education and preuniversity education. The circuits contained at least a toggle switch, a lamp, a battery, a voltmeter, and an ammeter. These components were supplemented in varying ways by other toggle switches, lamps, batteries, voltmeters, ammeters, and/or push switches, resistors, variable resistors, fuses, buzzers, and gears driven by constant speed motors.

Each circuit contained multiple faults; for example, batteries or meters could be connected in the wrong way, components could be short-circuited by redundant wire, or the voltage or current could be too high or too low. The participants were informed that good functioning, that is, after repair, encompassed the following: (a) all components with outwardly observable functions function visibly when the circuit is closed (e.g., lamps burn and gears turn visibly); (b) the (repaired) circuit contains at least the same components, that is, components could be added to the initial, malfunctioning circuit or could be changed but not removed; (c) all components are properly connected; and (d) in case of multiple switches, the circuit functions appropriately at different switch (on–off) combinations.

The tasks were preceded by a general introduction that was intended to familiarize participants with the functioning of Crocodile Physics. It contained a textual explanation of the program functions, an example of a simple circuit that participants had to copy so that they could practice with placing and connecting components, and an example of a simple troubleshooting task. The introduction also contained a simple task to practice concurrent reporting that was somewhat simpler than the actual experimental tasks, but not so simple that participants could solve it right away. So, they had time to familiarize themselves with thinking aloud, without being distracted by task demands. Furthermore, it contained a demonstration of an eye-movement recording (the cue), which showed the experimenter's eye movements and mouse keyboard actions on the same task that was used to practice concurrent reporting. Figure 1 gives an indication of what the cue looked like, by fictitiously displaying the replay of a record over one of the tasks in the cued retrospective reporting condition. The definition of a properly functioning circuit (as described above) was given halfway through as well as at the end of the introduction. (Note: Parts of the three preceding paragraphs describe information also present in van Gog et al., 2005, p. 210.)

Verbal instructions. Instructions and prompts were worded in line with the standards described by Ericsson and Simon (1993). The instruction before the practice task in the introduction was "you should really think aloud, that is, verbalize *everything* that comes to mind, and not mind my presence in doing so, even when curse words come to mind for example, these should also be verbalized. Act as if you were alone, with no one listening, and just keep talking." Whenever participants stopped verbalizing their thoughts, the experimenter prompted them after 5 s by saying, "Please try to keep talking." For the concurrent reporting condition, the instruction was "please think aloud while you are working on the next two tasks." The instructions to the retrospective and cued retrospective reporting conditions consisted of two parts. For the retrospective condition, the first part was "Please complete this task; you can work in silence." The second part was "This is the begin state of the task. Can you please tell me what you were thinking during problem solving?" The first part of the instruction to the cued retrospective condition was "I am going to record your eye movements while you are working on the next task, so I will calibrate the eye-tracking system before you start. In a minute you will see a red square appearing on your screen; please follow it with your eyes."

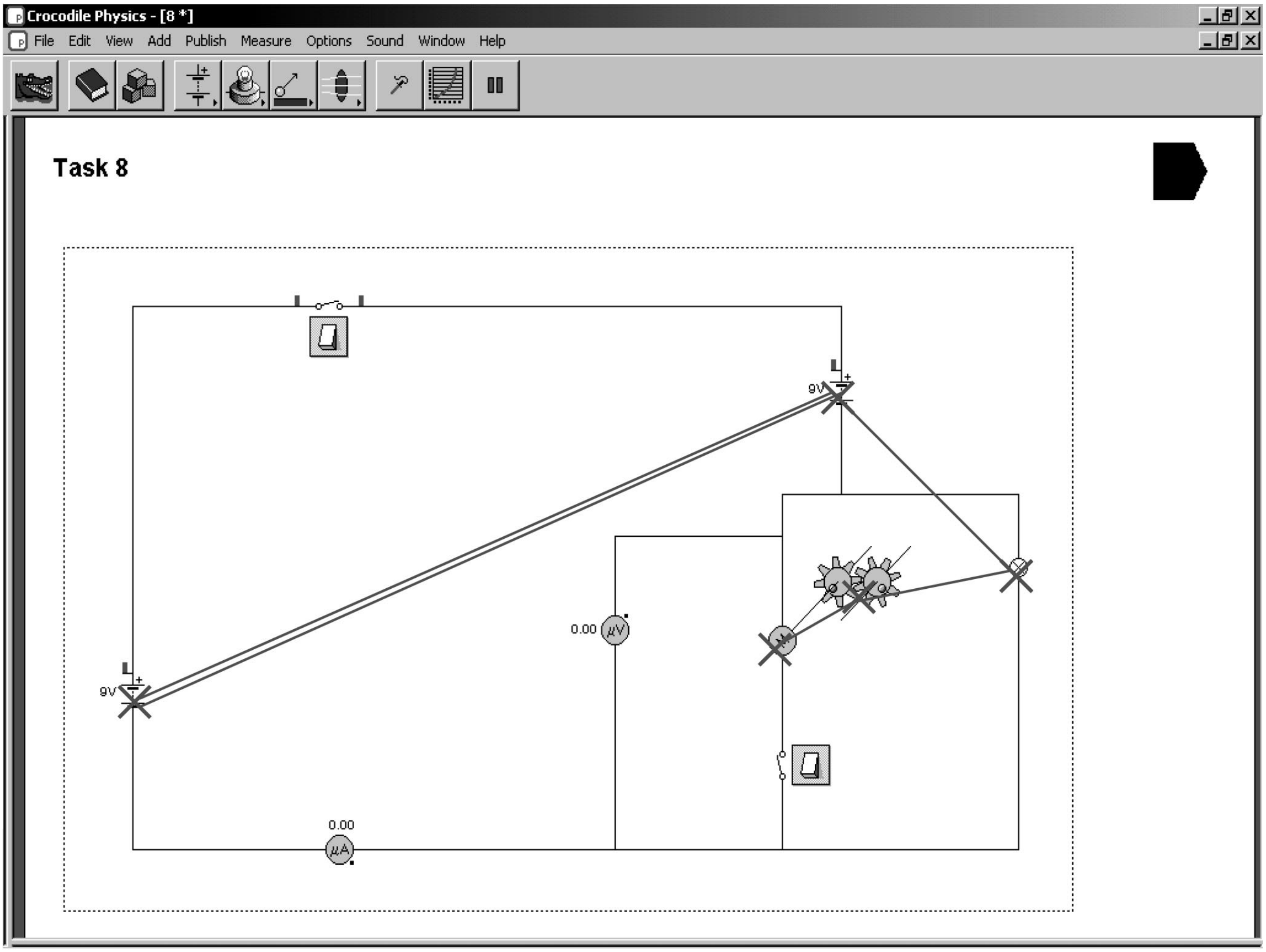

Figure 1. An example of the cue, showing the eye movements replayed over the task (replay of mouse–keyboard actions is not simulated here). Participants saw a red cross, which indicated their eye fixations, moving across the screen. This is fictitiously displayed here through multiple crosses connected by the thinner lines.

After calibration: "Thank you, the system is calibrated. Please complete the next task; you can work in silence." The second part of the instruction was "This is a record of your eye movements and your actions. I am going to replay it; please watch it and tell me what you were thinking during problem solving. If you want the record to pause, you can use the key F2; when you want it to proceed, you can press the 'play' button in the program menu." When participants stopped verbalizing, the prompting procedure described above was used in each condition.

Procedure

Each participant was scheduled for an individual session of approximately 90 min. Participants were seated in the chair in front of the stimulus PC in the recording room, and they were informed that they would start with an introduction to the program before they could start working on the tasks. The experimenter also indicated that the tasks would have to be completed under different conditions and that instructions regarding those conditions would be given at the right moment. If participants had no further questions, they were asked to put their head in the forehead rest and to adjust the chair so that they would be seated comfortably in that position when eye movements would be recorded (when eye movement recording was not necessary, they did not have to keep their head in the forehead rest). The experimenter then went into the observation room, and participants started with the introduction.

After finishing the introduction practice task, the participants were asked whether they felt comfortable enough with the concurrent reporting procedure to do it "for real," and the experimenter also judged whether this was the case before continuing. Although some participants initially needed quite a number of prompts, none of the participants or the experimenter deemed additional practice tasks necessary.

When they had finished the introduction, participants were informed that they were allowed to ask questions but that questions related to the content of the tasks would not be answered, only those with regard to the functioning of Crocodile Physics. Then they were given the appropriate instruction for the first task of their assigned sequence. The order of the instructions depended on the sequence of conditions to which participants were assigned, and instructions were given after they finished either one (in the retrospective and cued retrospective conditions) or both (in the concurrent condition) tasks in a condition. In the retrospective and cued retrospective conditions, the first part of the instructions was given before each task, and the second part was given after each task. Before each task in the cued retrospective condition, the eye-tracking system was calibrated, and the recording of eye movements and mouse–keyboard actions was started with the GazeTracker software. In the concurrent and retrospective

Table 1
The First, Median, and Third Quartile Values for the Number of Codes Assigned to Each Information Type Per Reporting Method

	Reporting method								
	Concurrent			Retrospective			Cued retrospective		
Information type	1st	Median	3rd	1st	Median	3rd	1st	Median	3rd
Action	17.50	39.00	65.75	12.00	16.00	22.25	19.00	28.50	36.75
Why	8.75	12.00	18.25	7.00	8.00	11.25	6.00	11.00	14.25
How	2.00	3.00	5.25	0.00	1.00	1.25	1.00	2.50	4.25
Metacognitive	0.75	3.00	6.25	1.00	2.00	3.00	2.75	5.50	7.00

conditions, only participants' mouse–keyboard actions were recorded with the GazeTracker software, which was started before both tasks (in the concurrent condition) or before each task (in the retrospective condition). After each task, a screen followed that indicated that they had to wait for the experimenter (in the retrospective and cued retrospective conditions) or for instructions. After retrospective reporting or cued retrospective reporting, the instructions for the next condition were given face to face by the experimenter. The concurrent and retrospective verbalizations were recorded on digital audiocassettes.

Data Analysis

Coding scheme. The coding scheme was developed on the basis of our definitions of information types and was refined by analyzing samples of the protocols. The task-oriented sub- and main categories, which were used in the later analyses, are shown in the Appendix. Next to these, categories like "questions asked," "answers given," "program-related remarks," and "no code" (i.e., unintelligible) were included. Protocols were segmented at a small grain size: Each sentence or utterance preceded and followed by a pause was considered a separate segment. If a segment clearly contained two types of information, it received two codes. Two raters were familiarized with the experimental tasks and program, so that they could meaningfully interpret possible references to the task or program in the protocols. They scored 20% of the protocols, which were not used in the refining phase of the coding scheme, with an interrater reliability of .79 (Cohen's kappa). When considering only the task-oriented main categories, the interrater reliability was also .79 (Cohen's kappa). Because the interrater reliability was sufficiently high (i.e., higher than .70; van Someren et al., 1994), one rater scored the remaining protocols.

Dependent variables. For each participant, the number of codes on the task-oriented main categories, action, why, how, and metacognitive information per method, was calculated by summing the number of codes on the constituting subcategories (see Appendix).

Results

The data were analyzed by using Friedman's tests with Conover's (1999) procedure for comparisons (an equivalent to the parametric Fischer's least significant difference procedure), with a .05 significance level (two-way). Table 1 shows the differences per method in the median values of the number of codes assigned to each category of information. Table 2 shows the sums of ranks. The Conover procedure is based on the difference between the sums of ranks. A difference is significant when the outcome of the following formula (Conover, 1999, p. 371) is correct (i.e., when the result in the left-hand part is indeed larger than the result in the right-hand part):

$$|R_j - R_i| \rangle t_{1-\alpha/2}\left[\frac{2(bA_1 - \sum R_j^2)}{(b-1)(k-1)}\right]^{\frac{1}{2}}$$

The number of codes on action information differed significantly between the three methods, $\chi^2(2, N = 26) = 12.51, p < .01$. The Conover procedure showed that in order to be significant at the .05 level, the difference between sums of ranks should exceed 12.56. As can be inferred from Table 2, for action information, the difference between concurrent protocols and retrospective protocols was 24.44, and the difference between retrospective and cued retrospective protocols was −16.12; hence, both were significant.[3] The direction of those differences can be inferred from Table 1. In line with our hypothesis, the number of codes on action information in concurrent protocols was higher than in retrospective protocols and was higher in cued retrospective protocols than in retrospective protocols.

Significant differences between the methods were also found in the number of codes on why information, $\chi^2(2, N = 26) = 7.15, p < .05$; how information, $\chi^2(2, N = 26) = 15.49, p < .01$; and metacognitive information, $\chi^2(2, N = 26) = 12.51, p < .01$. The Conover procedure showed that in order to be significant at the .05 level, the difference between sums of ranks should exceed 13.52 for why information, 11.06 for how information, and 12.05 for metacognitive information. As can be inferred from Table 2, on why and how information, there was a significant difference between concurrent and retrospective protocols (differences in sums of ranks of 19.24 and 23.66, respectively). However, Table 1 reveals that this effect was in the direction opposite of the predicted one: Concurrent protocols contained more why and how information than retrospective protocols. On metacognitive information, there was no significant difference (the difference in sums of ranks was 5.20).

Given this unexpected finding that retrospective protocols did not contain more why, how, and metacognitive information than concurrent protocols, it did not make much sense to test our initial hypothesis that cued retrospective protocols—like retrospective protocols—would contain more why, how, and metacognitive information than concurrent protocols. We therefore tested whether cued retrospective protocols—like concurrent protocols—also contain more why and how information and not significantly

[3] Please note that the comparisons reported in this section are not independent statistical analyses, as degrees of freedom have been exhausted by the previous analyses. Because it is not possible to compute confidence intervals for these effects, we report the traditional statistical test results, with the qualification that they are partially redundant with the other analyses.

Table 2
Sums of Ranks for Each Information Type Per Reporting Method

	Σ ranks			
Reporting method	Action	Why	How	Meta
Concurrent	62.92	61.62	61.10	47.58
Retrospective	38.48	42.38	37.44	42.38
Cued retrospective	54.60	52.00	57.46	66.04

Note. Meta = metacognitive.

different or more metacognitive information than retrospective protocols.

From Table 2 it can be inferred that there was no significant difference in the number of codes in why information in cued retrospective and retrospective protocols (the difference in sums of ranks was −9.62), but there was a significant difference in how and metacognitive information (differences in sums of ranks of −20.02 and −23.66, respectively). Table 1 reveals that cued retrospective protocols contained a higher number of codes on how and metacognitive information than retrospective protocols.

Discussion

Regarding our hypothesis that concurrent reporting would provide more action information than retrospective reporting and that retrospective reporting would yield more why, how, and metacognitive information than concurrent reporting, only the first part was confirmed. Unexpectedly, concurrent reporting not only resulted in more action information but also in more why and how information than retrospective reporting, and there was no significant difference on the amount of metacognitive information provided by these methods. A possible explanation for this unexpected finding may lie in the fact that we first segmented the protocols on the basis of utterances and then coded the segments, whereas Taylor and Dionne (2000) seemed to have coded meaningful episodes (in which case the segmenting and coding processes become intertwined).

Our hypothesis that cued retrospective reporting—like concurrent reporting—would elicit more action information than retrospective reporting was confirmed. Given the unexpected finding that concurrent reporting yielded a higher amount of why and how and a not significantly different amount of metacognitive information than retrospective reporting, we tested whether cued retrospective reporting would also result in more why, how, and metacognitive information than retrospective reporting. This was indeed the case for the how and metacognitive information, but there was no significant difference between cued retrospective reporting and retrospective reporting in the amount of why information elicited.

Some critical observations must be made with regard to this study. First of all, retrospective reports are known to be sensitive to fabrication. On the one hand, in cued retrospective reporting, fabrication of thoughts related to actions may be less likely because of the cue. On the other hand, the cue might lead to an active reconstruction of thoughts as a result of reviewing one's own problem-solving process, instead of being solely based on memory. On the basis of the present study, it is not possible to distinguish memory only and active reconstruction processes. This issue should be addressed in future research, especially because it might be an explanation for the fact that cued retrospective reporting did result in more metacognitive information being reported than retrospective reporting, whereas concurrent reporting did not. Second, we decided not to use probes in retrospective and cued retrospective reporting, but only prompts, and we kept instructions and prompts as similar as possible across conditions to avoid possible bias. Although this approach is less of a threat to reliability and validity, the use of probes is often seen as a great benefit of retrospective reporting and does not have to influence reliability and validity when probes are neutral, nonevaluative, and do not restrict the participant's reporting (Ericsson & Simon, 1993; Taylor & Dionne, 2000; van Someren et al., 1994). This implies that the full potential of both retrospective reporting and cued retrospective reporting may not have been realized in this study. Third, some technical aspects in cued retrospective reporting were not entirely optimal. Because of the real-time replay of the record, the speed was equal to that of the actual problem-solving session. It might be interesting to investigate whether and how slower replay or participant control over replay would affect the results.

In sum, both concurrent reporting (with the exception of metacognitive information) and cued retrospective reporting (with the exception of why information) resulted in a higher number of codes on the different types of information than did retrospective reporting. These results suggest that the method of cued retrospective reporting based on a record of eye movements and mouse–keyboard operations superimposed on the task may be a promising one for eliciting a broad range of information about solving computer-based problems. It seems worthwhile to further investigate the possible benefits of this method in relation to concurrent reporting. Some important qualitative questions should be addressed, such as whether concurrent and cued retrospective reports capture the same content and whether the methods are differentially effective for different groups of participants (e.g., different levels of expertise). Another qualitative issue that would be interesting to study is whether the different communicative nature of concurrent and cued retrospective reporting (e.g., the latter might be perceived as more conversational) influences the content of the protocols. If this is the case, it would be another methodological issue to address in future studies in addition to the ones described above.

References

Camps, J. (2003). Concurrent and retrospective verbal reports as tools to better understand the role of attention in second language tasks. *International Journal of Applied Linguistics, 13,* 201–221.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies, 41,* 801–849.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Hansen, J. P. (1991). The use of eye mark recordings to support verbal retrospection in software testing. *Acta Psychologica, 76,* 31–49.

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology, 113,* 387–404.

Lankford, C. (2000). GazeTracker: Software designed to facilitate eye movement analysis. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 57–63). New York: ACM Press.

Lauwereyns, J., & d'Ydewalle, G. (1996). Knowledge acquisition in poetry criticism: The expert's eye movements as an information tool. *International Journal of Human-Computer Studies, 45,* 1–18.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21,* 1–29.

Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 167–187). Mahwah, NJ: Erlbaum.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, England: Cambridge University Press.

Russo, J. E. (1979). A software system for the collection of retrospective protocols prompted by eye fixations. *Behavior Research Methods & Instrumentation, 11,* 177–179.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17,* 759–769.

Taylor, K. L., & Dionne, J. P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology, 92,* 413–425.

van den Haak, M., De Jong, M. D. T., & Schellens, P. J. (2003). Retrospective versus concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology, 22,* 339–351.

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science, 32,* 83–98.

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19,* 205–221.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modeling cognitive processes*. London: Academic Press.

(*Appendix follows*)

Appendix

The Task-Oriented Categories of the Coding Scheme

Category			
Sub	Main	Remarks relating to . . .	Example
1. Survey	Action	the organization of the circuit	"This one has three lamps, two connected in parallel and one in series."
2. Practical evaluation of initial state	Action	a practical evaluation of the initial state	"Nothing happens."
3. Evaluation of information in initial state	Action	evaluation of information from the Crocodile function, or components in the initial state	"The current here is 12 Volt."
4. Try	Action	trying the circuit	"Let's see what it does."
5. Execute	Action	actions/operations	"I am now adding a lamp."
6. Practical evaluation of changed state	Action	a practical evaluation of the executed solution step	"Damn, the lamp still explodes."
7. Evaluation of information in changed state	Action	evaluation of information from the Crocodile function, meters, or wires	"The lamp got 12 Volt, while its maximum is 9 Volt."
8. Theoretical evaluation of initial state	Why	a theoretical evaluation of the initial state	"The voltage is very low; the lamp won't burn."
9. Theoretical evaluation of executed solution step	Why	a theoretical evaluation of an executed solution step	"Now it should work."
10. Theoretical evaluation of possible solution step	Why	a theoretical evaluation of a possible but not executed solution step	"But if I lower the voltage, the other lamp gets too little."
11. Theory about circuits in general	Why	theory about circuits in general	"In parallel connections each branch gets the same voltage."
12. Theory about specific parts of circuits	Why	theory about components of circuits	"An ammeter should be connected in series."
13. Definition of problem	Why	the problem	"This battery is connected in the opposite direction from the others."
14. Definition of solution	Why	a (possible) solution	"The resistance should be higher."
15. Exclusion of possibilities	Why	exclusion of possible causes	"The batteries are properly connected, so that isn't the problem."
16. Evaluation of previous problem states	How	evaluation of previous problem states	"When the resistor was set at less Ohm, this lamp exploded."
17. Heuristics	How	the use of heuristics	"I always check first whether all components are connected properly."
18. General approach to/searching for a solution	How	a general approach to or searching for a solution	"How can I do this?"
19. Specific approach to a solution	How	a specific approach to a solution	"Then I am going to check this first."
20. Goal orientation	Meta	goal orientation	"All lamps should burn visibly."
21. Self-evaluation/knowledge	Meta	the evaluation of knowledge	"Remarkable how much I remember."
22. Self-evaluation/actions	Meta	the evaluation of actions	"This was not a smart thing to do."
23. Self-evaluation/strategy	Meta	the evaluation of strategy	"I am going in the right direction."
24. Task evaluation	Meta	the evaluation of the task	"This one is easy."

Note. Meta = metacognitive.

Received December 8, 2004
Revision received August 15, 2005
Accepted August 15, 2005 ■