Metacognition Learning (2008) 3:39–58
DOI 10.1007/s11409-007-9009-6

Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?

Maria Bannert · Christoph Mengelkamp

Received: 3 August 2005 / Accepted: 17 October 2007 /
Published online: 9 November 2007
© Springer Science + Business Media, LLC 2007

Abstract Recent research on metacognition points out the crucial role of on-line methods when endeavouring to conduct valid assessments of metacognitive skills. Presently, different on-line methods are used, however, it is still a question of research whether and how they affect students' learning behaviour and learning outcome. Thus, the aim of this study is to quasi-experimentally analyse the effects of two on-line verbalisation methods on learning performance. By means of the thinking-aloud method, students in one experimental group ($n=24$) were instructed to read and think aloud during learning. With the reflection when prompted method, students of another experimental group ($n=24$) were prompted at each navigational step to reflect on the reasons why they chose specific information. Students in the control group ($n=22$) learned without being instructed to verbalise. All three groups were treated identically except for the different use of verbalisation assessment methods. The students' task was to learn the concepts and principles of operant conditioning presented in a hypermedium within 30 min. The students' learning sessions were videotaped and learning performance was obtained immediately afterwards. Based on Ericsson and Simon's (*Protocol analysis: Verbal reports as data*, MIT, Cambridge, 1993) model, no performance differences between the thinking aloud and the control group were hypothesised. However, prompting students for metacognitive reflection should affect learning performance positively, which is confirmed by the results only in tendency for transfer performance. Implications for on-line assessment methods of metacognitive skill will be discussed.

Keywords Metacognitive assessment · Verbalisation · Thinking aloud · Reflection prompts · Hypermedia learning · Knowledge acquisition

M. Bannert (✉)
Faculty of Philosophy, Educational Media, Chemnitz University of Technology, Reichenhainertsr. 41, 09126 Chemnitz, Germany
e-mail: maria.bannert@phil.tu-chemnitz.de

C. Mengelkamp
Center of Educational Research, University of Koblenz–Landau, Landau, Germany

Introduction and research aim

This paper addresses the assessment of metacognitive skills with two different on-line methods: thinking aloud and reflection when prompted. The question of possible method effects on learning process and performance was raised while conducting a comprehensive research project on metacognitive support for hypermedia learning (Bannert 2004, 2007). Because the design and evaluation of adequate metacognitive support should be based on an appropriate identification of metacognitive skills, valid metacognitive assessment methods are needed urgently. The problem of valid metacognitive skill diagnosis has often been neglected in metacognitive research, however, this question has recently been put on the research agenda by some researchers (e.g. Artelt 2000; Dominowski 1998; Lompscher 1996; Pintrich et al. 2000; Pressley 2000; Schraw and Impara 2000; Veenman 2005; Winne and Perry 2000).

In the following, the general aim of the research project will be described briefly, and a self-developed model of metacognitive skills necessary for hypermedia learning will be outlined. Then, different on-line metacognitive skill assessment methods will be sketched with a focus on the thinking aloud procedure. On the basis of the model developed by Ericsson and Simon (1993), three different levels of verbalisation and their effects on cognitive processes and performance will be described, which will lead to the research questions and hypotheses of the current study.

Metacognitive skills during hypermedia learning

Recent research in the field of hypermedia learning points out the crucial role of learners' strategic and metacognitive behaviour (e.g. Foltz 1996). Although metacognitive knowledge and skills are needed when learning without using new learning technology, this technology makes the students' reflective behaviour regarding their own way of learning more salient (e.g. Lawless and Brown 1997; Lin et al. 1999; Rouet and Levonen 1996). Based on linear text processing research, Schnotz (1998) points out that higher metacognitive demands of non-linear text, respectively hypertext, compared with traditional linear text is partly caused by the fact that in a non-linear learning environment, a successful learner has to continuously decide where to go next and to constantly evaluate how the information retrieved is related to his or her actual learning goal. In a conventional linear text, however, the author guides the reader through the learning topic. Moreover, a hypermedia environment requires further strategic decisions by learners as the student has to choose permanently not only between various text nodes, but also between distinct information presentation formats (e.g. information codality: description vs. depiction, cf. Schnotz and Bannert 2003, and information modality: visual vs. auditory information (see Low and Sweller 2005). Although this assumption of higher metacognitive demand of hypermedia environments is often postulated, comprehensive empirical validation is still in progress. Despite this gap in research between the metacognitive demands of hypertext versus traditional text, it is well founded that many students have difficulties in performing appropriate strategic and metacognitive learning behaviour (e.g. Pressley et al. 1987, 1989; Simons and De Jong 1992). Thus, the general aim of the research project is to provide suitable scaffolding for metacognitive reflection when learning with hypermedia. Though strongly related, the aim of this paper is clearly different as it focuses not on support, but rather on assessment issues of metacognition.

In recent research on metacognition, a distinction between metacognitive knowledge and metacognitive skills has been made (e.g. Hasselhorn 1992; Schraw 2001). On the one hand, *metacognitive knowledge* refers to the individual's declarative knowledge of learning strategies, person and task characteristics which are relevant mastering a specific situation (Flavell and Wellman 1977). On the other hand, *metacognitive skills* refer to the control, monitoring, and self-regulation activities that take place when learning and solving problems (Brown 1978). In our research, the focus lies on the second component of metacognition, i.e. the student's metacognitive skills during hypermedia learning.

According to relevant research, it is assumed that successful learning with hypermedia or web-based environments is not a matter of trial and error, but rather a set and specific sequence of metacognitive activities which have to be performed (e.g. Astleitner 1997; Schnotz 1998). This research suggests that a successful learner performs different metacognitive activities when learning with hypermedia, as visualised in Fig. 1. Therein, a self-developed model is presented which summarises existing assumptions and models of metacognitive skills and serves as the main theoretical framework for this study. One has to keep in mind that at the current state of the art, the model is hypothetical rather than empirically validated and, thus, should be interpreted as a proposal for future hypermedia learning research. In addition, as it is based on learning and metacognition with traditional linear text (Pressley 2000; Pressley and Afflerbach 1995; Schnotz 1998) it includes processes which are also performed in other learning contexts. According to this hypothetical model, a successful learner analyses the situation first before he or she starts with the execution of information processing. He or she will *orientate* him- or herself by glancing over task, instruction, and resources ("What is there to do?," "What resources do I have?"), specify the *learning goals* or even break them down into sub-goals ("What should I know afterwards?") and *plan* the ongoing procedure ("Which sequence?," "How much time is available?"). Based on this analysis, the student has to *search* for the relevant information and—especially crucial for web-based learning—*judge* whether the information found is really relevant to reaching the learning goals. He or she then has to extract the information (it may not be necessary to process all the information presented in one link)

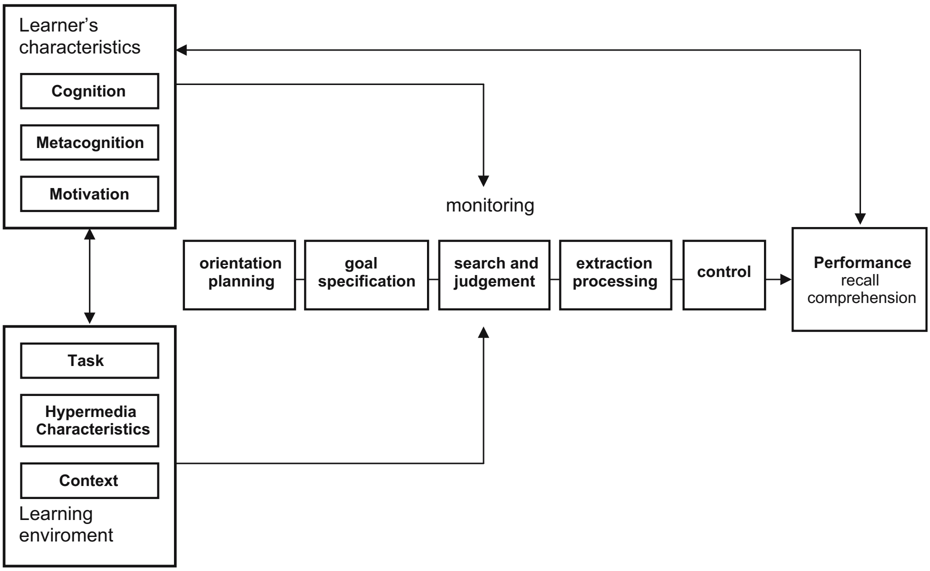

Fig. 1 Model of metacognitive activities during hypermedia learning

and to *elaborate* on it (deep processing). After learning, he or she has to *evaluate* the learning outcome, again with respect to the learning goals ("Have I achieved all learning goals?," "Did I comprehend the subject matter?"). These activities are constantly *monitored* and *controlled*.

This hypothetical learning process is determined by the *learner's characteristics*, such as prior knowledge, metacognitive knowledge and skills, cognitive ability (Pressley et al. 1987, 1989), and also motivation, which is highlighted in current research on the one hand (e.g. Boekaerts 1997; Schneider 1996). On the other hand, *external conditions* also play a crucial role, for example, the task characteristics or the specifics of the hypermedia system (hierarchical structure, existence of a navigational menu or map, guided tour, glossary etc.) and the context (Astleitner 1997).

Recent research on hypermedia learning reveals that many learners have difficulties in performing these metacognitive activities spontaneously, which most probably results in lower learning outcomes (Balcytiene 1999; Bannert 2003; Brenstein and Schellhas 1998; Hill and Hannafin 1997; Lawless and Brown 1997; McManus 2000; Rouet and Levonen 1996). Recently, efforts have been made to develop effective metacognitive instructions urgently demanded by other researchers (e.g. Jacobson et al. 1996; Kramarski and Feldman 2000). However, in order to evaluate the metacognitive instruction's effectiveness, one needs valid assessment methods to identify whether the metacognitive skills that have just been explicated are really performed. The empirical studies cited above use different methods to assess metacognitive strategies, mainly questionnaires (McManus 2000) and interviews (Balcytiene 1999; Brenstein and Schellhas 1998). The explorative study by Hill and Hannafin (1997) employs think aloud protocols to qualitatively assess the use of learning and metacognitive strategies. Due to the huge effort of scoring verbal protocols, data of only four participants are reported in their study. In the next section, think aloud as one method to assess the use of metacognitive skills is described in more detail.

Verbal data as on-line assessment of metacognition

In his recent review, Veenman (2005) classified metacognitive skill assessment methods as off-line and on-line methods. Depending on the moment they are conducted, he distinguishes *off-line methods*, which are performed prospectively or retrospectively to learning, from *on-line methods*, which are conducted concurrently during learning. *Prospective* and *retrospective* assessments are usually obtained by questionnaires and interviews and often fail to predict learning outcome. Therefore, for assessing metacognitive skills in particular, he recommends on-line methods, which are more accurate and valid. In the following, we use the definitions of off-line and on-line as proposed by Veenman. Beyond the methods discussed by Veenman (2005), judgements of knowledge are used in research on metamemory and metacomprehension. Pintrich et al. (2000) include these judgements of knowledge in the term 'on-line,' whereas Desoete et al. (2003) use the term 'off-line' for these judgements. For a systematic review of pros and cons of assessment methods available, see Veenman (2005), Winne and Perry (2000), and Pintrich et al. (2000).

When assessing metacognitive skills during hypermedia learning, *concurrent assessment* could be conducted with systematic observations via logfile or eye movement analyses or verbalisation methods. Among these, verbalisation methods are the only ones that allow a glimpse into the participant's mind during learning, whereas data from eye movements or logfiles concern behaviour and, thus, the cognitions and metacognitions underlying the behaviour have to be interpreted by the researcher. For example, a category system is built to obtain an indicator of deep orientation activities, indicated by the amount of time a subject spends on the content table or the index (Beishuizen and Stoutjesdijk 1999). Thus, the data obtained are interpreted as the amount of metacognitive orientation activity. In contrast, data from think aloud protocols gives insight into the mind of the person whilst the content table or the index are on the screen, and can thus be categorised as metacognitive orientation involving less interpretation than data from logfiles. Data from the recording of eye movements lie somewhere in between, because they indicate what part of the screen the person is looking at, but, similarly to logfiles, the data render no information about what comes to the mind of the person while looking at the information presented on that part of the screen. Because data from logfile and eye movement methods need more interpretation than data from think aloud protocols with respect to metacognitive activities (e.g. Beishuizen and Stoutjesdijk 1999; Prins 2002; Veenman 1993), the focus in this study lies on the on-line verbalisation method of thinking aloud.

Thinking aloud method

When the thinking aloud method is employed, participants are asked to talk aloud during thinking, problem solving, and/or learning, and these verbal protocols are analysed by means of coding schemes (Afflerbach 2000; Chi 1997; Ericsson and Simon 1993; Pressley and Afflerbach 1995; van Someren et al. 1994). The aim is to identify cognitive and metacognitive processes underlying task performance in different subject areas and contexts. Although this method is frequently used in psychological and educational research, it has been criticised from the very beginning till today (see Afflerbach 2000; Crutcher 1994; Nisbett and Wilson 1977; Payne 1994; Russo et al. 1989). To summarise, this criticism mainly concerns two problems. The first one refers to the question of the *reactivity* of the method: Is the process of thinking altered throughout the method of think aloud because thinking aloud needs resources of the cognitive system that could otherwise be used for the primary task? The second problem refers to the *completeness* of the verbal protocols: Are the protocols obtained by think aloud procedures complete, or is any information about the cognitive processes missing? The model by Ericsson and Simon described very briefly in the next section is helpful when dealing with these two questions.

Ericsson and Simon's model

To answer the questions above, it is useful to distinguish between three different types of verbal data, as proposed in the model by Ericsson and Simon (1993, p. 16ff): talk aloud (level 1 verbalisation), think aloud (level 2 verbalisation), and verbalisation procedures that involve mediating processes before verbalisation (level 3 verbalisation), e.g. introspection or reflection. In short, these methods differ in the ways the verbal reports are obtained (see the section below). The model by Ericsson and Simon is based on models of human cognitive architecture which assume that human knowledge is stored in different buffers (Anderson 2004): In the *sensory register* (SR), the information is encoded into an internal form and is stored very briefly. Most information is lost quickly, only a small amount of information enters the *short term memory* (STM) by means of attention. The STM is the active and conscious part of the memory system and is limited in capacity and duration. It holds information only for a short time. Information is bumped out by new information or fades away. By means of rehearsal or elaboration processes, information will attain to the *long-term memory* (LTM). The LTM permanently stores all information which is not immediately used. To process this information, it has to be retrieved into the STM. According to this cognitive model, only the conscious content of the STM/working memory can be verbalised. The content of the LTM, however, can not be verbalised directly, but first has to be retrieved back into the STM. Furthermore, the automatic processes can not be verbalised.

With regard to different kinds of intermediate processes between information access and verbalisation, Ericsson and Simon (1993) distinguished three types of verbal data mentioned briefly before. They are described in more detail in the following because they require different types of instructions and, furthermore, they affect the underlying cognitive processes in a different way. Consequently, both aspects have to be considered carefully when choosing the assessment method, regardless of whether it will be applied in a nonlinear or linear learning environment.

- *Level 1 verbalisation*: At this level, the participant says the verbally encoded content of his or her STM out loud without any specific effort or intermediate processes. Thus, the conduction of the primary task is not affected at all; there are neither changes in cognitive processes, nor performance effects. Ericsson and Simon (1993) called this first verbalisation level *Talk Aloud*.
- *Level 2 verbalisation*: At this level, mediation processes take place as the STM content has to be verbally encoded in order to explicate it because the content is not encoded in a verbal form originally, but e.g. in a figural one. This verbal encoding or explication requires processing time, leading to longer time for the task and probably for short pauses whilst speaking. However, according to Ericsson and Simon, it does not affect the structure of the cognitive processes for performing the primary task. Ericsson and Simon called this second verbalisation level *Think Aloud*.
- *Level 3 verbalisation*: At this level, mediation processes are necessary as the participant has to explain his or her thoughts, ideas, or motives. These additional interpretative processes require not only more processing time, but much more importantly, may change the cognitive processes compared to cognitive processes without verbalisation. Information from the LTM has to be retrieved and linked with the content of the STM in order to draw inferences. Additionally, decisions as to whether the information in the STM should be verbalised or not have to be made with so-called filtering processes. Since this verbalisation is "not simply a recoding of content already present in STM, but requires linking this information to earlier thoughts and information attended to previously" (Ericsson and Simon 1993, p. 79), it will affect the primary task performance.

In order to obtain these different verbal reports, one has to formulate the *verbalisation instruction* with care. For level 1 and level 2 verbalisations, students have to be instructed to say every thought out loud which comes into their mind during task performance without any interpretation, explanation, or judgement. In case of silence, the participant is prompted by the experimenter "to continue thinking aloud" or "to keep talking." In comparison, for level 3 verbalisations, students have to be instructed to reflect, which can be formulated differently, depending on the specific aim of the study. For example, a participant could be instructed to say out loud why he or she considered certain information or to explain how he or she solved the problem. Often, this important formulation difference is mentioned without caution in hypermedia research and, thus, methodology artefacts may occur (e.g. Davidson-Shivers et al. 1997, 1999). We would like to point out here that according to the model by Ericsson and Simon aiming at the identification of spontaneous use of metacognitive skills, only level 1 and 2 verbalisation instructions are adequate on-line assessment methods, because all instructions requiring reflection and/ or filtering lead to level 3 verbalisations. Consequentially, the major importance of Ericsson and Simons's model with respect to metacognitive on-line assessment is that the (meta)cognitive processes are altered throughout the specific verbalisation method used. To test this assumption empirically, one part of the participants in this study was asked about reflections at certain times while they performed a task. Subsequently, we will refer to this method as *reflect when prompted*. Table 1 summarises the main evidence and implications for learning performance on these three verbalisation levels.

The problem of consciousness

Although talk aloud, think aloud, and reflect when prompted require a certain amount of the participants' verbal ability, most participants learn easily to say their thoughts out loud and, respectively, their reflections within a few minutes (Ericsson and Simon 1993, see also Pressley and Afflerbach 1995). However, one has to keep in mind that according to the model by Ericsson and Simon, only conscious activities can be verbalised. Therefore, unconscious processes taking place during learning can not be assessed by these verbal methods. In the case of assessing metacognition, this is not really a problem for those researchers who claim consciousness to be a major characteristic of metacognition (e.g. Garner 1988; Paris et al. 1983). However, other researchers argue that metacognitive processes may not be conscious or statable in many learning situations (e.g. Brown 1978; Reder 1996) because they are often highly automated or developed without any conscious reflection. Recently, most researchers' definitions lie in between (e.g. Baker 1994; Schneider and Weinert 1990; Schnotz 1992; Veenman 1993): Metacognition is conscious *or* at least accessible to consciousness when difficulties during task performance such as comprehension problems or errors occur.

Empirical findings of studies investigating different verbalization methods

In some studies, the effects of verbalisation methods on learning process and learning outcome were analysed empirically (for an overview, see Ericsson and Simon 1993; van Someren et al. 1994), however, without explicit focus on hypermedia learning. In their review, Ericsson and Simon (1993, p. 106) conclude: "When the instructional procedures conformed to our notion of Level 1 or Level 2 verbalization, the studies gave no evidence that verbalization changes the course or structure of the thought processes." De Jong (1987) assessed metacognitive skills by using two different thinking aloud methods. One of them involved participants being instructed to think aloud constantly, whereas the other one involved them being prompted to think aloud only at marked points. In conclusion, data produced by both on-line methods predicted learning outcomes much more accurately than scales of the off-line questionnaire. Unfortunately, statistics of verbalisations obtained by

Table 1 Levels of verbalisations and their effects on learning performance

Three levels of verbalisation	Effects
1. Talk aloud → No intermediate process	No performance effects
2. Think aloud Verbal encoding process → Longer processing time	Time, but no performance effects
3. Reflect when prompted Inference, filtering, verbal encoding → Changes cognitive processes	Time and performance effects

means of the instruction to think aloud constantly vs. markedly were not reported. Nevertheless, De Jong recommended the instruction method of thinking aloud constantly for assessing metacognitive skills. Students often feel more disturbed when instructed with marked prompts compared with constant thinking aloud, leading to less acceptance (see also Drewniak 1992). In a similar study, Crain-Thoreson et al. (1997) analysed the effects of constant thinking aloud vs. marked thinking aloud vs. reading aloud. In conclusion, these three methods did not procure different learning outcomes. In Veenman's (1993) experiment, the participants of the group that thought aloud continuously reached similar learning scores compared with the participants of the control group that learned silently. However, participants of the thinking aloud group needed more time to perform the learning task. These results correspond with the assumptions of level 2 verbalisations as stated by Ericsson and Simon (1993).

However, there are also empirical studies that have demonstrated an influence of the thinking aloud procedure on cognitive processing and task performance. For example, thinking aloud altered the performance in choosing gambles and computing simple additions, but not in solving anagrams and Raven's matrices (Russo et al. 1989). Russo and colleagues considered thinking aloud a secondary task which taps additional cognitive resources which, in turn, would decrease primary task performance. Moreover, Schooler et al. (1993) explained the lower problem solving scores which they had obtained in the thinking aloud group with interference processes. They argued that the thinking aloud procedure verbally overshadows the insights necessary to solve problems, especially when solving insight problems.

Because the effects of on-line metacognitive assessment methods on learning processes and metacognitive activities are still a question of research, some authors recommend controlling their effects by using a control group design, i.e. including a group of participants learning without verbalisation (e.g. Russo et al. 1989). Furthermore, other researchers recommend multiple method designs by using different assessment methods simultaneously (e.g. Beishuizen and Stoutjesdijk 1999; Veenman 2005).

Reflect when prompted as a tool for metacognitive support

To sum up the main theoretical and empirical evidence, methods using level 3 verbalisation (e.g. reflect when prompted) are not recommended for assessing student's spontaneous use of metacognitive skills because of their reactivity to the (meta-) cognitive processes and, thus, possible method artefacts. However, in metacognitive training, the method of reflection prompts seems to be an adequate intervention as it prompts students to reflect upon their learning process and to apply their metacognitive skills. The prompting focuses students' attention on their own thoughts and on understanding the activities they are engaged in during the course of learning. Hence, it is assumed that prompting students to reflect upon their own way of learning will allow them to activate their repertoire of metacognitive knowledge and skills, which will further enhance hypermedia learning and transfer (e.g. Lin et al. 1999; Lin and Lehman 1999).

Research questions and hypotheses

The aim of this study is to experimentally analyse the effects of two on-line verbalisation methods of metacognitive skills assessment on learning performance during hypermedia learning. Furthermore, the method's effects on retrospective judgements of learning process and perceived disorientation during learning are investigated. Based on the research sketched above—especially the distinction of three different verbalisation levels postulated in the model by Ericsson and Simon (1993)—this study will mainly address two research questions.

The first question is whether the on-line method of metacognitive skills assessment will influence *learning performance*. According to the model by Ericsson and Simon (1993), it is hypothesised that learning performance is influenced by the type of verbalisation method used. Because level 3 verbalisations obtained with instructions to reflect when prompted may change the (meta-) cognitive processes, this on-line method should significantly alter learning performance. Further, theory of metacognition in hypermedia learning suggests that the processes are altered in a manner that increases the learning outcome via a higher metacognitive awareness. Generally, according to Ericsson and Simon's model, there should be no differences in process between the thinking aloud group and the control group with no thinking aloud instruction. However, as thinking aloud requires more processing time, less learning outcome may occur when both groups learn within the same fixed time interval. In short, the learning performance of the reflection when prompted group should be significantly better compared with the performance of the control group and the thinking aloud group. There may be performance differences between the last two groups due to time constraints.

The second research question of this study refers to the effects of the on-line method of metacognitive skills assessment with respect to the *learning process*, student's *perceived interruption* and *perceived disorientation* during hypermedia learning. According to research in metacognition and hypermedia learning, it is assumed that metacognitive and strategic activities during learning will increase when students are prompted to reflect on metacognitive and strategic aspects during learning (e.g. Lin et al. 1999; Lin and Lehman 1999; Simons and De Jong 1992; van den Boom et al. 2004; Veenman 1993). This assumption is also supported by Ericsson and Simon's (1993) model which—as presented above—postulates (meta-)cognitive processes altered by level 3 verbalisations, respectively reflection when prompted verbalisations. Thus, with respect to the *learning process*, it is hypothesised that the reflect when prompted group should report upon significantly more strategic and metacognitive activities immediately after learning than the thinking aloud and the control groups. In addition, student's *perceived disorientation* should be decreased in comparison to the thinking aloud and control groups because the higher metacognitive awareness induced by the reflect when prompted instruction should correspond with a better orientation. Nevertheless, participants in the reflect when prompted group should judge this verbalisation method as more *interrupting*. Although participants usually become easily automised with the thinking aloud procedure, students in the thinking aloud group should also report upon more interruptions during learning than the control group which was not instructed to talk or think aloud.

Method

Sample and design

The effects of metacognitive assessment methods were analysed by means of a quasi-experimental study. To investigate the research questions, data from 70 university students participating in two different studies, counterbalanced according to prior knowledge, metacognitive knowledge, verbal intelligence, and motivation, were re-analysed. Table 2 presents the sample descriptions and confirms the samples' comparability with the aid of inference statistics.

The original aim of Study 1 was to investigate the relationship between students' metacognitive skills used spontaneously during hypermedia learning, learning process, and learning performance (Bannert 2005; Mengelkamp 2001). Metacognitive activities were subjected to a multiple measurement by means of the on-line thinking aloud method as well as with an off-line retrospective questionnaire. Students in this group, referred to as the *thinking aloud group* ($n1=24$), were instructed to follow the on-line level 2 verbalisation method postulated by Ericsson and Simon (1993), i.e. to read and think aloud during hypermedia learning without being asked to give any explanation or interpretation.

The primary aim of Study 2 was to experimentally investigate the effects of reflection prompts with respect to hypermedia learning process and learning performance (Bannert 2006; Weinand 2003). Students in the group referred to as the *reflection when prompted group* ($n2=24$) were prompted at each navigation step in the hypermedia system to name the reasons why they had chosen this specific information node out loud, whereas the *control group* ($n3=22$) learned silently, i.e. without reflection prompting and without thinking aloud.

The two studies differ only with respect to the different on-line assessment instructions described above (Study 1: thinking aloud—level 2 verbalisations, Study 2: reflect when prompted—level 3 verbalisations vs. control group—no verbalisations). Students participating in study 2 were allocated randomly to the reflect when prompted and the control group. The data listed in Table 2 statistically support the samples' comparability in variables that may potentially have an influence on the learning outcome and/or on the metacognitive processes during learning. Since learning procedure and materials, learning instructions, learning time, instruments, etc. were identical, data of both studies were re-analysed quasi-experimentally with regard to group differences arising from different on-line verbalisation methods. Thus, a total data of 70 undergraduate university students majoring in different fields (mean age=24.20, SD=4.88; female: 84.3%) were analysed. As mentioned above, the three verbalisation groups did not differ regarding learner characteristics.

Table 2 Sample descriptions

	Study 1	Study 2		$F_{(2,67)}$	p
	Think aloud (n=24) *M* (SD)	Reflection when prompted (n=24) *M* (SD)	Control group (n=22) *M* (SD)		
Age	24.04 (4.23)	24.17 (4.78)	24.41 (5.80)	0.032	0.968
Semester	4.04 (2.60)	3.00 (2.89)	3.91 (2.84)	0.965	0.386
Prior knowledge	7.83 (3.82)	7.58 (4.64)	7.23 (3.61)	0.129	0.880
Verbal intelligence	26.58 (5.43)	27.38 (4.92)	27.32 (3.08)	0.216	0.806
Motivation for achievement	7.46 (2.60)	7.17 (2.46)	8.64 (2.92)	1.939	0.152
Fear of failure	8.88 (3.86)	8.92 (3.88)	7.86 (3.81)	0.543	0.584
Metacognitive and Strategic Activities					
Orientation, planning	3.71 (0.83)	3.79 (0.70)	3.73 (0.83)	0.055	0.946
Monitoring/ regulation	4.83 (0.64)	4.72 (0.76)	4.50 (0.56)	1.504	0.230
Organisation	4.72 (0.99)	4.50 (0.80)	4.56 (0.76)	0.414	0.663
Elaboration	4.49 (0.82)	4.44 (0.86)	4.33 (0.89)	0.207	0.813
Rehearsal	2.52 (0.67)	2.49 (0.76)	2.66 (0.87)	0.318	0.728

Procedure and material

As visualised in Fig. 2, learner's characteristics were obtained by questionnaire about 1 week before the participants started learning. Prior knowledge was measured with a self-developed multiple-choice test, metacognitive knowledge with a modified version of the LIST questionnaire (Wild et al. 1992), verbal intelligence with the IST 2000 (Amthauer et al. 1999), and motivation for achievement and fear of failure with the LMT (Hermans et al. 1978). The LIST questionnaire is normally used to assess the use of learning strategies during the course of university studies in general. The modification mainly consisted in describing a specific learning setting regarding which the items have to be answered. Before the students filled in the questionnaire, they were instructed to imagine learning a complex text for an examination under time pressure conditions and answering all items in regard to their assumed behaviour in such a situation. This was similar to the situation the students were in 1 week later when the hypermedia learning session took place. Because of this modification, some LIST-scales (e.g. learning with fellow students or searching for literature) were no longer suitable and, thus, were eliminated from the questionnaire. This modification did not affect the factor structure and the scales' reliabilities (see Bannert 2007).

Each study began with an *introduction phase*. First, students were shown how to navigate in the hypermedia programme which took place by means of html-scripts using the Netscape explorer. Afterwards, they were introduced to the method of reading and thinking aloud (EG1) or to the method of reflect when prompted (EG2). In order to practice verbalising, they had to carry out several search tasks using a similar programme to the one in the learning session.

Following this, the *learning session* began. Students had to learn specific concepts and principles of operant conditioning within a fixed time interval of 30 min. The *thinking aloud group* (EG1) had to read and think aloud constantly. In case of silence, the participant was prompted by the experimenter to continue reading and thinking aloud. The participants in the *reflect when prompted group* (EG2) were prompted to reflect on their navigation steps. That is, with every navigation step, respectively node selection, they were prompted to name the reasons why they had chosen this specific information node. In case of silence,

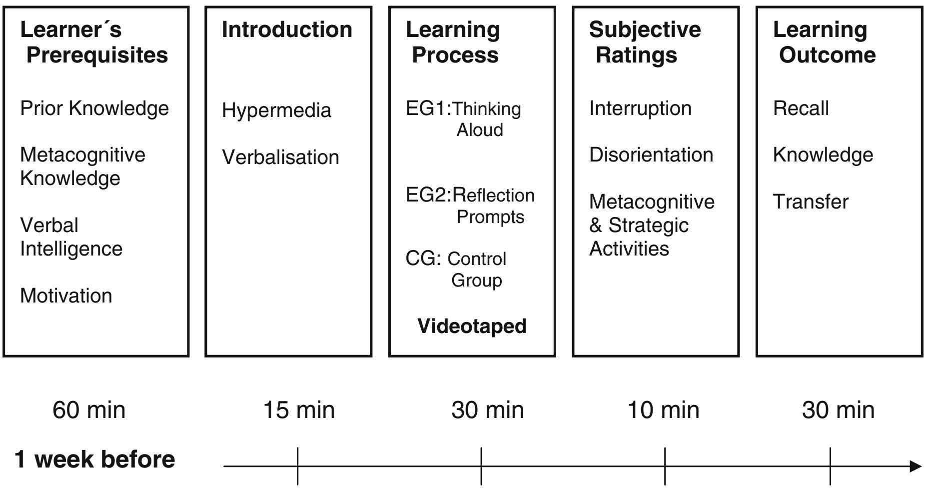

Fig. 2 Design and procedure of the studies

the participant was prompted by the experimenter to name his or her reasons for node selection out loud. Participants of the *control group* (CG) learned silently without any on-line verbalisation method instructions. The students were completely free in navigation. Each learning session was videotaped.

In the *testing phase* which took place immediately after learning, students first had to judge interruptions perceived during learning and then to fill in a questionnaire about perceived disorientation. The disorientation questionnaire contained the translated items of the Scale of Perceived Disorientation developed by Beaseley and Waugh (1995). Afterwards, learning outcome was measured with self-constructed tests. Finally, strategic learning behavior was assessed retrospectively by using another modified version of the LIST questionnaire (Wild et al. 1992). This time, the items were re-formulated to assess the learning process that had ended just before retrospectively. Scales concerning time management, effort, and concentration were eliminated because they are not suited to assess learning strategies during learning for 30 min retrospectively, but rather global strategies or learner characteristics. The studies were conducted in individual learning sessions and each took about 1 1/2 h.

The *hypermedia learning environment* consisted of 44 nodes with about 12,500 words, 19 pictures/ diagrams, and 240 links in total. The part to be learned involved 9 nodes, including 2,300 words, 3 pictures and 60 links. Navigation was made possible by using the hierarchical navigation menu, the forward- and backwards-buttons on each node and the hotwords directly placed in the text.

Results

Instruments

In Table 3, descriptions of the questionnaires' scales used for further analyses are presented. *Learning Performance* was measured with paper–pencil-tests. *Free recall* was operationalised by counting the basic terms and concepts students wrote down on a blank sheet of paper (max=55). *Knowledge* attained was measured with a multiple-choice test including 22 items, each with one correct and three false alternatives, and as a fifth alternative the option "do not know." A correct answer was rewarded with one point (max=22 points, α=0.70). *Transfer* was measured by asking students to apply the basic concepts and principles they had just learned to solve prototypical problems in educational settings, such as how a teacher should behave when classroom discipline becomes problematic. Answers were rated based on a self-developed rating scheme (max=40 points, interrater agreement—Kappa=0.84, α=0.69).

The *scale of disorientation* including ten translated items of the Scale of Perceived Disorientation developed by Beaseley and Waugh (1995) reached high reliability (α=0.83).

Perceived interruptions were measured with one item, asking students in the reflect when prompted group whether they felt interrupted by this method. Students in the thinking aloud group had to judge their perceived interruptions arising from the thinking aloud procedure, and participants in the control group were asked to judge whether they felt disturbed by the video camera which was present in all three treatment groups. This item can not assess the different reasons, but only the degree of perceived interruptions.

As described above, students judged their *strategic activities* during hypermedia learning retrospectively by means of the modified LIST scales (Wild et al. 1992). Factor structure was obtained with the same procedure (principal component analysis with varimax rotation) and conducted separately for metacognitive and cognitive strategies. Items with high factor

Table 3 Statistical descriptions of retrospective questionnaires used

Scale	No. of items	Cronbach's alpha	Item means	Item variance
Learning performance				
Free recall	max 55	–	–	–
Knowledge	22	0.70	0.64	0.19
Transfer	8	0.69	2.34	1.18
Strategic activities				
Orientation, planning	8	0.69	3.16	1.85
Monitoring/regulation	6	0.79	4.37	1.46
Rehearsal	2	0.65	3.29	1.82
Organisation	5	0.78	3.53	3.18
Elaboration	7	0.77	3.34	1.97
Perceived disorientation	10	0.83	2.65	1.41
Perceived interruptions	1	–	3.80	1.53

Recall: max=55, Knowledge: max=22, Transfer: max=40, Strategic activities, perceived disorientation, perceived interruptions: min=1, max=6

loadings on more than one factor and items with low discrimination indices were not included in the final version of the scales. Table 3 shows the scales, all with good or at least sufficient reliabilities. Metacognitive strategies were measured with two scales: *Orientation and Planning* (α=0.76), and *Monitoring and Regulation* (α=0.84). These two factors explain 45.1% of variance. For cognitive strategies, three scales were constructed explaining 57.0% of variance: *Organisation* (α=0.81), *Elaboration* (α=0.83), and *Rehearsal* (α=0.76). Reliabilities of the cognitive scales are equal to or even higher than those of the LIST (Wild et al. 1992; Schreiber 1998).

Learning performance

It was assumed that students learning with reflection prompts would achieve better learning outcomes compared with the control and thinking aloud groups. At the end of the studies, learning performance was measured by means of recall, knowledge and transfer tasks.

Table 4 presents the mean performance scores for each treatment group. Although the reflection when prompted group achieved higher knowledge and transfer performance, these differences are not significant, as supported by the results of Anovas. In addition, post-hoc comparisons based on the Tukey-HSD revealed no significant group differences. Nevertheless, a one-tailed t test shows a significant effect between the reflect when prompted and control groups for transfer performance, $t(1,44)=3.640$, $p<0.05$, $d=0.55$. Thus, there is a tendency in favour of the reflect when prompted group.

Thus, the variation of the verbalisations method of metacognitive skills assessment had no significant effect on recall, knowledge, or on transfer task performance and, therefore, our hypotheses on performance effects due to the type of verbalisation method have to be rejected.

Metacognitive and strategic activities observed with the off-line retrospective questionnaire

It was hypothesised that prompting for reflection would increase metacognitive and strategic behaviour during hypermedia learning. To test this assumption, the scales of the retrospective questionnaires concerning metacognitive and strategic activities conducted during learning were analysed with respect to group differences.

Table 4 Means and standard deviations of students' learning performance, metacognitive and strategic activities, perceived disorientation and interruption

Measure	Reflect when prompted (n=24)		Think aloud (n=24)		Control group (n=22)		F(2,67)	p
	M	SD	M	SD	M	SD		
Learning performance								
Recall[g]	11.54	4.23	14.08	6.87	12.14	5.39	1.34	0.268
Knowledge[h]	14.17	4.05	13.63	3.52	13.95	3.18	0.14	0.873
Transfer[i]	20.21	4.73	18.48	5.58	17.77	3.84	1.59	0.212
Perceived disorientation[j]	2.67	0.77	2.77	0.71	2.40	0.71	1.48	0.235
Perceived interruptions[j]	4.08[a]	1.50	3.75[b]	1.39	2.32[a,b]	1.46	9.47	0.000
Metacognitive and strategic activities[j]								
Orientation, planning	3.18	0.72	3.09	0.94	3.22	0.62	0.15	0.864
Monitoring/regulation	4.31	0.98	4.56	0.80	4.23	0.73	1.01	0.369
Rehearsal	2.67[c]	1.11	3.88[c]	1.08	3.34	0.98	7.84	0.001
Organisation	2.76[d,e]	1.37	3.95[d]	0.99	3.90[e]	1.22	7.45	0.001
Elaboration	2.92[f]	0.87	3.64[f]	0.92	3.47	0.79	4.52	0.014

Significant group differences according to Tukey-HSD post-hoc comparison with $p<0.05$ are indicated by letters [a] to [f]

[g] max=55

[h] max=22

[i] max=40

[j] min=1, max=6

Table 4 contains the means of each scale calculated for each treatment group. Whereas there are no differences regarding the metacognitive scales Orientation/Planning and Monitoring/Regulation, scores for the cognitive learning scales clearly differ. Students learning with reflection prompts retrospectively reported less strategic activities. According to their ratings, they conducted less Rehearsal, Organisation, and Elaboration activities, whereas students in the thinking aloud group judged their learning behaviour with respect to these three activities as significantly more strategic, as shown by the results of Tukey-HSD post hoc comparisons. In general, our hypotheses about different effects of on-line methods on strategic learning behaviour had to be rejected since group differences do not point in the expected direction. However, one has to keep in mind that these results are based on students' subjective ratings. Below, they will be correlated with more objective process data.

Perceived disorientation and perceived interruptions observed with the off-line retrospective questionnaire

Immediately after learning, students were asked to answer a questionnaire measuring perceived disorientation and interruptions during learning. As one can see in Table 4, ratings for perceived disorientation during hypermedia learning do not differ significantly in the treatment groups. Thus, our hypothesis about the influence of verbalisation methods on perceived disorientation has to be rejected.

Perceived interruptions were measured with one item asking students to rate whether they felt interrupted by the assessment method (i.e. EG1: thinking aloud, EG2: reflection prompts, CG: video camera). As expected, participants of the reflection when prompted group judged this method as more interrupting (s. Table 4). Furthermore, according to a Tukey-HSD post-hoc comparison, participants of the thinking aloud procedure also perceived more interruptions during learning than the control group, although significantly lower in comparison to the prompting group, as expected. Thus, our hypothesis about the degree of perceived interruptions has been confirmed by the data. However, due to item construction, no statements can be made about the respective reasons which may differ greatly between the three treatment groups.

Construct validity: correlations between questionnaire and observational data

With the following explorative analysis, we investigated the construct validity of the two different metacognitive assessment methods we applied in the present study. Students' ratings of metacognitive and strategic learning activities measured retrospectively by means of an *off-line questionnaire* (s. scales in Table 3) were correlated with the activities obtained by the video analysis and the data regarding transfer performance.

For *the reflect when prompted group* (EG2, n=24) all *reasons expressed for node selections* were rated by means of a self-developed coding scheme, the categories of which were mainly deductively constructed based on the theoretical framework presented at the beginning (in detail see Bannert 2006). First, the statements were transcribed, and then their quality was rated by means of the coding scheme. All coding was conducted by a trained rater and took about 3 h per participant. In case of ambiguous statements, a second trained rater was consulted to conduct the final coding. For each participant, a global score for adequate strategic and metacognitive reasons was calculated by adding all strategic statements together. On average, 17 statements (M=16.6, SD=6.66) were coded as strategic reasons for node selection. This score correlates significantly with transfer performance (r=0.49, p=0.015), but not with any scale of the strategic learning questionnaire (listed in Table 3).

Similarly, *video protocols of the thinking aloud group* (EG1, n=24) were analysed by means of a different coding scheme also based on the theoretical framework (in detail see Bannert 2004). Here, the video-protocols were analysed with respect to which quality the students actually performed different activities in, visualised in Fig. 1 (Orientation & Planning, Goal Specification, Search & Judgement, Processing, Control, Monitoring). Maximum scoring for each strategic activity was two, and minimum scoring was zero. For example, when students articulated and performed planning activities by reflecting what to do next, and even wrote down the plan, they received two points, and no points when they did not show this activity during learning at all. One point was given to those who did not perform planning activities such as these in the optimal manner. Interrater agreement of two independent raters was Kappa 0.79. For each participant, a global score for adequate metacognitive activities was calculated by means of all six coded activities (M=0.90; SD=0.50). This score also correlates significantly with transfer performance (r=0.56, p=0.005), but not with any scales of the questionnaires except the elaboration scale (r=0.54, p=0.006).

In a nutshell, questionnaire scales do not correlate with the learning performance data. Thus, our observational data is a much better predictor for learning performance and, moreover, does not correspond with the scales of the retrospective questionnaire (except for the elaboration scale in the thinking aloud group).

Discussion

In regard to the assumptions derived from the research on metacognition (e.g. Pressley et al. 1987, 1989; Schraw 2001; Veenman 2005), hypermedia learning (e.g. Bannert 2007; Lin and Lehman 1999; Schnotz 1998), and the thinking aloud methods in particular sketched above (Ericsson and Simon 1993; Pressley and Afflerbach 1995; van Someren et al. 1994), our hypotheses were only partly confirmed by the results. Although participants who were instructed with reflection prompts achieved a marginally better transfer performance, varying the on-line verbalisation assessment method had no significant effect on learning outcomes. In line with Ericsson and Simon's (1993) approach, no performance differences were obtained for the thinking aloud and control groups, even though learning times were fixed for all treatment groups and, therefore, some effect could have occurred because level 2 verbalisation slows down cognitive processes slightly. However, the thinking aloud procedure conducted during hypermedia learning did not decrease the learning outcome. This result offers valuable clues for future research. Often, the thinking aloud method is not used as a metacognitive assessment method due to the question of its underlying reactivity (Russo et al. 1989; Schooler et al. 1993). Although we can not firmly conclude from the results of this study that thinking aloud does not affect metacognitive processes, we can, however, definitely conclude that thinking aloud does not affect learning performances relatively to the control condition. We interpret this as indirect evidence that thinking aloud does not interfere with metacognition and, therefore, we recommend this type of verbalisation as a sensitive metacognitive on-line assessment method.

In the present study, the students were not prompted to reflect on certain aspects, but could give any reason for card selection. Presumably, reflection when prompted requiring metacognitive reasons and statements would have shown an increase in the expected learning performance. However, this speculation has to be investigated in a follow-up study based on the same methodological approach. In this context, the question of the suitable point of time to prompt successfully should also be investigated. The amount of navigational steps, respectively card selections, which have to be reflected on within a specific period of time could presumably have a negative impact on learning motivation and hinder strategic learning. As research is lacking, we want to experimentally manipulate the time- or behaviour-based interval for prompting in order to be able to give recommendations for more adaptive learner-centred prompting.

As expected, participants in the reflection when prompted group judged this method as more interrupting. This result corresponds with other studies in some respects, with participants being prompted to think aloud not continuously, but rather at certain times, such as every 5 min (e.g. Drewniak 1992) or after reading a few sentences (De Jong 1987). In these two studies, students felt particularly disturbed in the case of this so-called time-marked prompting. Although one has to keep in mind that the comparison between these studies is not really valid because our study used the prompted to reflect verbalisation method, whereas the other two studies used the prompted to think aloud verbalisation method, we decided to prompt students at each new card selection in order to reduce this feeling of interruption. We assumed this method would disturb the reading and learning process less compared with fixed time interval-prompting. On average, participants in the reflect when prompted group selected one node selection per minute which had to be accompanied by a reflection statement. Again, this is also incomparable to Drewniak's procedure of prompting every 5 min. What is more, we also assumed that our prompting method would interfere with students' learning according to the model by Ericsson and Simon (1993), which is confirmed by our data.

However, perceived disorientation during hypermedia learning was not affected by the verbalisation method. Effects of the reflection when prompted method on metacognitive and strategic activities even went in the opposite direction to what we had expected. Participants in the reflection when prompted group reported less strategic activities in the questionnaire, especially compared with the participants in the thinking aloud group. At first glance, this could be interpreted as a strong concentration on card selection which each had to be reflected on by making at least one statement, which may impede on strategic learning activities. However, one has to keep in mind that learning performance—which requires strategic activities—is marginally better in the reflection when prompted group. Indeed, to test this assumption, further research using additional on-line measures is required, such as logfile or eye-movement methods. Based on this kind of mixture of on-line methods, one could find out whether the reflect when prompted group actually perform less, and the thinking aloud group actually perform more strategic learning activities.

Finally, in accordance with Veenman's (2005) recent review, there was no significant correlation between the scales of the questionnaires and activities obtained by the protocol analysis. Moreover, no significant correlation was obtained for self-report data and learning performance. One interpretation could be that the questionnaire data reflects the amount of metacognitive and cognitive strategic activities quite well, and therefore is a valid instrument with respect to *quantity* of strategic behaviour. But we hypothesise that *quality* of strategic behaviour during learning is not measured adequately with questionnaires. For example, a student may report accurately that she monitored a lot during learning, but may not be aware of the low quality of her monitoring which may be taking place in inadequate phases of learning or may lead to her judging something as understood that was, in fact, misunderstood, etc. In contrast, the data obtained by think-aloud and reflect when prompted methods indicate the quality of metacognitive strategic activities, as these data are obtained by judging the appropriateness of the strategy with respect to the hypothetical model of “ideal learning” described at the beginning of the paper. This assumption would explain why there is no correlation between the questionnaire data and the data obtained by verbalisation methods (excepting elaboration in the think-aloud group). Further, this hypothesis explains why there is no significant correlation between questionnaire data and learning performance, but a correlation between data stemming from verbalisation methods and transfer performance. For future research, the assumption that questionnaires are indicators of quantity of metacognition, whereas verbalisation methods allow coding of the quality of metacognition, has to be reconsidered.

Thus, although it is rather time-consuming, it is necessary to include such on-line process analyses in further research on metacognitive skills assessments and in-depth evaluation of metacognitive support. According to the results of the present study, the thinking aloud methods based on level 1 and level 2 verbalisations do not seem to significantly affect meta-/cognitive processes during hypermedia learning. Our interpretation of comparable learning performance as indirect evidence for the non-reactivity of thinking aloud with metacognition certainly has to be investigated in a more direct manner.

In a broader sense, this research suggests activating student's metacognitive skills during hypermedia learning (e.g. Bannert 2004, 2006). Prompting methods, such as reflection prompts, seem to be effective instructional measures, especially for students who already possess these skills, but who do not perform them spontaneously (so-called production deficit, e.g. Hasselhorn 1995). However, for students lacking metacognitive competence (so-called mediation deficit), extensive training is necessary, which could also be accompanied by metacognitive prompting. Regardless of which kind of metacognitive instruction is adequate in respect to target groups and specific learning contexts, valid methods are needed to assess the student's metacognitive activities during learning. Based on this study, we recommend the talk and thinking aloud method.

Acknowledgement This study was partially supported by the DFG [Deutsche Forschungsgemeinschaft, BA2044/1].

References

Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. L. M. Kamil, P. B. Mosenthal, D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 163–179). Hillsdale, NJ: Erlbaum.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). *IST 2000—Intelligenz-Struktur-Test 2000 [Intelligence Structure Test 2000]*. Göttingen: Hogrefe.

Anderson, J. R. (2004). *Cognitive psychology and its implications.* New York: Worth.

Artelt, C. (2000). *Strategisches Lernen [Strategic Learning]*. Münster: Waxmann.

Astleitner, H. (1997). *Lernen in Informationsnetzen. Theoretische Aspekte und empirische Analysen des Umgangs mit neuen Informationstechnologien aus erziehungswissenschaftlicher Perspektive [Learning with information networks. Theoretical aspects and empirical analyses of dealing with new information technology from an educational perspective]*. Frankfurt: Europäischer Verlag der Wissenschaften.

Baker, L. (1994). Fostering metacognitive development. *Advances in Child Development and Behavior, 25*, 201–239.

Balcytiene, A. (1999). Exploring individual processes of knowledge construction with hypertext. *Instructional Science, 27*, 303–328.

Bannert, M. (2003). Effekte metakognitiver Lernhilfen auf den Wissenserwerb in vernetzten Lernumgebungen [Effects of metacognitive help on knowledge acquisition in net-based learning environments]. *Zeitschrift für Pädagogische Psychologie, 17*(1), 13–25.

Bannert, M. (2004). Designing metacognitive support for hypermedia learning. In H. Niegemann, D. Leutner, & R. Brünken (Eds.), *Instructional design for multimedia-learning* (pp. 19–30). Münster: Waxmann.

Bannert, M. (2005). Explorationsstudie zum spontanen metakognitiven Strategie-Einsatz in hypermedialen Lernumgebungen [Exploring the spontaneous use of metacognitive strategies during web-based learning]. In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 127–151). Münster: Waxmann.

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research, 4*, 359–375.

Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedien. [Metacognition in Hypermedia Learning]*. Münster: Waxmann.

Beaseley, R. E., & Waugh, M. L. (1995). Cognitive mapping architectures and hypermedia disorientation: An empirical study. *Journal of Educational Multimedia and Hypermedia, 4*, 239–255.

Beishuizen, J. J., & Stoutjesdijk, E. T. (1999). Study strategies in a computer assisted study environment. *Learning and Instruction, 9*, 281–301.

Boekaerts, M. (1997). Self-Regulated Learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*, 161–186.

Brenstein, E., & Schellhas, B. (1998). *Lernstrategien in hypermedialen Lernumgebungen* [Learning strategies in hypermedia learning environments]. Vortrag auf dem 39. Kolloquium zur Lern- und Lehrforschung. Universität Potsdam.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 77–165). Hillsdale, NJ: Erlbaum.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*, 271–315.

Crain-Thoreson, C., Lippman, M. Z., & McClendon-Magnuson, D. (1997). Windows on comprehension: Reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology, 89*(4), 579–591.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science, 5*(5), 241–244.

Davidson-Shivers, G. V., Rasmussen, K. L., & Bratton-Jeffery, M. F. (1997). Investigating learning strategies generation in a hypermedia environment using qualitative methods. *Journal of Computing in Childhood Education, 8*(2/3), 247–261.

Davidson-Shivers, G. V., Shorter, L., & Jordan, K., & Rasmussen, K. L. (1999). Learning strategies and navigation decisions of children using a hypermedia lesson. *Journal of Educational Multimedia and Hypermedia, 8*(2), 175–188.

De Jong, F. (1987). Differences in the self-regulation processes between successful and less successful students and the prediction of learning performances in case of comprehension and learning of text. In P. R. J. Simons & G. Beukhof (Eds.), *Regulation of learning* (pp. 33–45). Den Haag: SVO-Selecta.

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can off-line metacognition enhance mathematical problem solving? *Journal of Educational Psychology, 95*(1), 188–200.
Dominowski, R. L. (1998). Verbalization and problem solving. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 25–45). Mahwah, NJ: Erlbaum.
Drewniak, U. (1992). *Lernen mit Bildern in Texten. Untersuchung zur Optimierung des Lernerfolgs bei Benutzung computerpräsentierter Texte und Bilder [Learning with pictures in texts. Investigation to optimise learning success by using computer presented texts and pictures]*. Münster: Waxmann.
Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge: MIT.
Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail & W. Hagen (Eds.), *Perspectives on development of memory and cognition* (pp. 3–31). Hillsdale, NJ: Erlbaum.
Foltz, P. W. (1996). Comprehension, coherence, and strategies in hypertext and linear text. In J. F. Rouet, J. J. Levonen, A. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition* (pp. 109–136). Hillsdale, NJ: Erlbaum.
Garner, R. (1988). Verbal-report data on cognitive and metacognitive strategies. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies. Issues in assessment, instruction, and evaluation* (pp. 63–76). San Diego, CA: Academic.
Hasselhorn, M. (1992). Metakognition und Lernen [Metacognition and learning]. In G. Nold (Ed.), *Lernbedingungen und Lernstrategien. Welche Rolle spielen kognitive Verstehensstrukturen?* (pp. 35–63). Tübingen: Gunter Narr Verlag.
Hasselhorn, M. (1995). Kognitives Training: Grundlagen, Begrifflichkeiten und Desiderate [Cognitive training: Foundations, terms, and desiderates]. In W. Hager (Ed.), *Programme zur Förderung des Denkens bei Kindern* (pp. 14–40). Göttingen: Hogrefe.
Hermans, H., Petermann, F., & Zielinski, W. (1978). *LMT—Leistungsmotivationstest* [LMT—performance motivation test]. Amsterdam: Swets & Zeitlinger.
Hill, J. R., & Hannafin, M. J. (1997). Cognitive strategies and learning from the World Wide Web. *Educational Technology Research and Development, 45*(4), 37–64.
Jacobson, M. J., Maouri, C., Mishra, P., & Kolar, C. (1996). Learning with hypertext learning environments: Theory, design and research. *Journal of Educational Multimedia and Hypermedia, 5*(3/4), 239–281.
Kramarski, B., & Feldman, Y. (2000). Internet in the classroom: Effects on reading comprehension, motivation and metacognitive awareness. *Educational Media International, 37*(3), 149–155.
Lawless, K. A., & Brown, S. W. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science, 25*, 117–131.
Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36*(7), 837–858.
Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J.(1999). Designing technology to support reflection. *Educational Technology Research and Development, 47*(3), 43–62.
Lompscher, J. (1996). Erfassung von Lernstrategien auf der Reflexionsebene [Assessment of learning strategies on reflection level]. *Empirische Pädagogik. Zeitschrift zu Theorie und Praxis erziehungswissenschaftlicher Forschung, 10*(3), 245–276.
Low, B. R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147–158). Cambridge: Cambridge University Press.
McManus, T. F. (2000). Individualizing instruction in a web-based hypermedia learning environment: Nonlinearity, advanced organizers, and self regulated learners. *Journal of Interactive Learning Research, 11*, 219–251.
Mengelkamp, C. (2001). *Lautes Denken zur Erfassung von Lernstrategien* [Thinking aloud to assess learning strategies]. Unveröffentlichte Diplomarbeit, Fachbereich Psychologie, Universität Koblenz-Landau.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*, 293–316.
Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*(5), 241–248.
Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln, NE: Buros Institute of Mental Measurements.
Pressley, M. (2000). Development of grounded theories of complex cognitive processing: Exhaustive within- and between-study analyses of think-aloud data. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 262–296). Lincoln, NB: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.
Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. In R. Vasta & G. Whitehurst (Eds.), *Annals of child development* (Vol. 4, pp. 89–129). Greenwich, CT: JAI.

Pressley, M., Borkowski, J. G., & Schneider, W. (1989). Good information processing: What it is and how education can do to promote it. *International Journal of Educational Research, 13*, 857–867.

Prins, F. J. (2002). *Search & see. The roles of metacognitve skillfulness and intellectual ability during novice inductive learning in a complex computer-simulated environment.* Enschede, Niederlande: PrintPartners Ipskamp.

L. M. Reder (Ed.) (1996). *Implicit memory and metacognition*. Mahwah, NJ: Erlbaum.

Rouet, J. F., & Levonen, J. J. (1996). Studying and learning with hypertext: Empirical studies and their implications. In J. F. Rouet, J. J. Levonen, A. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition* (pp. 9–23). Hillsdale, NJ: Erlbaum.

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

Schneider, W. (1996). Zum Zusammenhang zwischen Metakognition und Motivation bei Lern- und Gedächtnisvorgängen [Coherence between metacognition and motivation in learning and memory processes]. In C. Spiel, U. Kastner-Koller, & P. Deimann (Eds.), *Motivation und Lernen aus der Perspektive lebenslanger Entwicklung* (pp. 121–133). Münster: Waxmann.

W. Schneider & F. E. Weinert (Eds.) (1990). *Interactions among aptitudes, strategies, and knowledge in cognitive performance*. New York: Springer.

Schnotz, W. (1992). Metacognition and self-regulation in text processing: Some comments. In M. Carretero, M. L. Pope, R. J. Simons, & J. I. Pozo (Eds.), *Learning & instruction. European research in an international context* (Vol. 3, pp. 365–375). Oxford: Pergamon.

Schnotz, W. (1998). Strategy-specific information access in knowledge acquisition from hypertext. In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools, and reasoning. Essays on situated cognition*. Heidelberg: Springer.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*, 141–156.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*, 166–183.

Schraw, G. (2001). Promoting general metacognitive awareness. In H. Hartman (Ed.), *Metacognition in learning and instruction. Theory, research and practice* (pp. 3–16). Dordrecht: Kluwer.

G. Schraw & J. C. Impara (Eds.) (2000). *Issues in the measurement of metacognition*. Lincoln, NE: Buros Institute of Mental Measurements.

Schreiber, B. (1998). *Selbstreguliertes Lernen. Entwicklung und Evaluation von Trainingsansätzen für Berufstätige [Self-regulated learning. Development and evaluation of training approaches for the employed]*. Münster: Waxmann.

Simons, P. R. J., & De Jong, F. P. (1992). Self-regulation and computer-assisted instruction. *Applied Psychology: An International Review, 41*, 333–346.

van den Boom, G., Paas, F., van Merrienboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior, 20*, 551–567.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method*. London: Academic.

Veenman, M. V. J. (1993). *Intellectual ability and metacognitive skill: Determinants of discovery learning in computerized learning environments*. Amsterdam: University of Amsterdam.

Veenman, M. V. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. S. 77–99). Berlin: Waxmann.

Weinand, C. (2003). *Metakognitive Förderung beim hypermedialem Lernen* [Metacognitive support on hypermedia learning]. Unveröffentlichte Diplomarbeit, Fachbereich Psychologie, Universität Koblenz-Landau.

Wild, K. P., Schiefele, U., & Winteler, A. (1992). *LIST. Ein Verfahren zur Erfassung von Lernstrategien im Studium [LIST. A scale for assessing learning strategies at studies]. (Gelbe Reihe: Arbeiten zur Empirischen Pädagogik und Pädagogischen Psychologie, Nr. 20)*. Neubiberg: Universität der Bundeswehr, Institut für Erziehungswissenschaft und Pädagogische Psychologie.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic.