

Learner Performance in Multimedia Learning Arrangements: An Analysis Across Instructional Approaches

Tessa H. S. Eysink

Ton de Jong

University of Twente

Kirsten Berthold

University of Freiburg

Bas Kolloffel

University of Twente

Maria Opfermann

Duisburg-Essen University

Pieter Wouters

Utrecht University

In this study, the authors compared four multimedia learning arrangements differing in instructional approach on effectiveness and efficiency for learning: (a) hypermedia learning, (b) observational learning, (c) self-explanation-based learning, and (d) inquiry learning. The approaches all advocate learners' active attitude toward the learning material but show differences in the specific learning processes they intend to foster. Learning results were measured on different types of knowledge: conceptual, intuitive, procedural, and situational. The outcomes show that the two approaches asking learners to generate (parts of) the subject matter (either by self-explanations or by conducting experiments) led to better performance on all types of knowledge. However, results also show that emphasis on generating subject matter by the learner resulted in less efficient learning.

KEYWORDS: instructional technologies, learning environments, constructivism

As the idea of active learning has become more prominent in educational science, a wide variety of new instructional methods or approaches has been developed that advocates learners' active attitude toward the learning material. This means that learners should try to make sense of the information offered by selecting relevant aspects of the presented material, organizing that relevant information into a coherent mental representation, and

integrating it with other information and prior knowledge (Mayer, 2001). At a more detailed level, active learning encompasses such processes as interpreting, exemplifying, classifying, inferring, differentiating, and organizing. It is assumed that meaningful learning occurs when learners engage in these cognitive processes (Mayer, 2002). As a consequence, instructional methods that elicit these processes are expected to be more successful in promoting meaningful learning than instructional methods that do not (Bransford, Brown, & Cocking, 2000; Merrill, 2002).

In studies in which new instructional methods that advocate active processing are compared, methods are often treated as extremes. For instance, in the often quoted paper by Kirschner, Sweller, and Clark (2006), unguided or minimally guided methods are judged against fully guided instruction. We would like to stress that in this article we are not treating approaches as caricatures. Instead, we adopt a more practical stance comparing realistic instructional approaches that were developed in such a way that they could also have been implemented in educational settings. More specifically, in the present article we compare four often used instructional strategies. From the learner's point of view, they involve: (a) Tell me how it works, (b) show me how it works, (c) let me explain how it works, (d) let me investigate how it works. The first strategy (tell me how it works) makes traditional expository instruction more active by presenting the same information in a variety

TESSA H. S. EYSINK is an assistant professor in the Department of Instructional Technology at the University of Twente, P.O. Box 217, 7500 AE, Enschede, the Netherlands; e-mail: *T.H.S.Eysink@utwente.nl*. Her research interests include inquiry learning, instructional design, and learning from multiple representations.

TON DE JONG is a full professor of educational technology at the University of Twente; e-mail: *A.J.M.dejong@utwente.nl*. His research interests are inquiry learning, technology enhanced learning, and science problem solving.

KIRSTEN BERTHOLD is an assistant professor in the group Educational and Developmental Psychology at the University of Freiburg; e-mail: *kirsten.berthold@psychologie.uni-freiburg.de*. Her research focus is on self-explanations, instructional explanations, and learning from multiple representations.

BAS KOLLOFFEL is a researcher at the University of Twente; e-mail: *B.J.Kolloffel@utwente.nl*. His research interests include effects of external representations and simulation-based inquiry learning.

MARIA OPFERMANN is an assistant professor at Duisburg-Essen University; e-mail: *maria.opfermann@uni-due.de*. She focuses on multimedia and cognitive load research as well as on research about the diagnosis and training of children's mathematical abilities.

PIETER WOUTERS is a researcher at Utrecht University; e-mail: *pieterw@cs.uu.nl*. His research focuses on cognitive and motivational processes in learning from computer games.

of different ways. The theory often identified as underlying this approach is cognitive flexibility theory, which states that presenting learning material from multiple perspectives leads to a flexible knowledge base (Spiro, Feltovich, Jacobson, & Coulson, 1992). The second strategy (show me how it works) originates from social learning theory (Bandura, 1976) and has found its way into cognitive apprenticeship (Collins, Brown, & Newman, 1989). In this approach, the expert's strategy is gradually laid out for the learners. The third strategy (let me explain how it works) centers around self-explanations (Chi, Bassok, Lewis, Reimann, & Glaser, 1989) and asks learners to reflect and self-explain principles from the learning material. The fourth strategy (let me investigate how it works) originates from the ideas of Bruner (1961) and Dewey (1938) and advocates that learners learn from performing investigations. The modern variant of this approach is guided inquiry learning in which scaffolds of all kinds structure the inquiry process (see, e.g., de Jong, 2006a).

Research on new instructional methods often focuses on comparing the effectiveness of the new approach to that of a more traditional method, for instance in large-scale evaluations (Cognition and Technology Group at Vanderbilt, 1992; Linn, Lee, Tinker, Husic, & Chiu, 2006). Though valuable in their own right, these comparisons have a number of disadvantages even when they use the same assessment methods for both the new and the traditional approach. First, when comparing a new approach with the existing curriculum it is very hard to keep the domain content the same for both conditions. Second, teachers as the medium of the instructional process have considerable influence on learner performance (see Brophy, 1986, for a review). The teacher's role is a variable that is difficult to control across approaches when comparing different instructional approaches.

We have overcome these two threats to the internal validity of comparison studies by using contemporary computer-based learning environments. Advances in computer technology have made it possible to design computer-based learning experiences that involve complex interactions between learners and instructional content (Reiser & Dempsey, 2002). The use of multiple representations in multimedia environments in combination with user interactivity allows the realization of new instructional approaches and their complementary cognitive tools and scaffolds in computer-based learning environments. Careful design of these learning environments can lay the foundation for a comparison of instructional approaches in which content, assessment, and procedure are held constant and in which the role of the teacher is eliminated.

In this study we have taken up this challenge, comparing four popular instructional approaches on learning effectiveness and efficiency. The four approaches that were selected for comparison follow from the four strategies described previously. The instructional approaches are: (a) hypermedia learning (tell me), (b) observational learning (show me), (c) self-explanation-based learning (let me explain), and (d) inquiry learning (let me investigate). The

study implemented the four instructional approaches in computer-based learning environments. The environments were designed using the latest insights in instructional design. All four computer environments focused on the domain of mathematics and shared the same content, namely, probability theory.

Instructional Approaches

Each of the four selected instructional approaches has its potentials for learning. This study compares the four approaches on different types of performance. We were interested in differences in effectiveness and efficiency and we also wanted to investigate whether different approaches result in different types of knowledge. This section gives a short overview of the four selected instructional approaches, including the potentials and claims of each. Table 1 summarizes this information by giving descriptions, assumptions underlying the approach, strengths and risks of the approach, and instructional variables that are of interest. Furthermore, the (sparse) research about learning with each of the four approaches in the domain of mathematics in general and probability theory in particular is discussed.

Hypermedia Learning: Tell Me

A hypermedia learning environment is a nonlinear computer environment in which multimedia components (e.g., text, pictures, animation, video) are stored in nodes that are interconnected by hyperlinks (Rouet, Levonen, Dillon, & Spiro, 1996). The multimedia components give access to information presented in different representational codes (e.g., verbal and pictorial) that may address different sensory modalities (auditory and visual) and that can be used in an interactive way. Due to their network-like information structure, hypermedia environments offer a variety of opportunities for self-regulated learning in addition to the presentation of information (Azevedo, 2005). Learners are free to decide which piece of information they want to select and observe and which they want to ignore. They can do this when they want to, at their own pace, and following their own order. Hypermedia learning environments can be characterized by a high level of interactivity, also referred to as a high degree of learner control. They are thought to lead to deeper understanding, as they allow for adaptive information retrieval and thereby enable active, flexible, and constructive learning (Spiro & Jehng, 1990). The risk of this flexibility, though, is that learners experience difficulties in selecting and integrating relevant information (Lawless & Brown, 1997; Rouet et al., 1996). Moreover, learners can experience structural and semantic disorientation (Hill & Hannafin, 1997). Instructional variables that are of interest in learning with hypermedia include different levels of learner control and support for representational and navigational choices (Burke, Etnier, & Sullivan, 1998). In the domain of probability theory, a study by Scheiter, Gerjets, and Catrambone (2006) showed that learning with hypermedia resulted in improved performance on near transfer tasks when learners

Table 1
Description, Assumptions, Strengths, Risks, and Instructional Variables of Interest for Each Instructional Approach

	Description	Assumptions	Strengths	Risks	Instructional Variables
Hypermedia learning	Fragments of information are stored in nodes that are interconnected and that can be accessed by electronic hyperlinks	Learners are able to identify their knowledge needs and select information accordingly Learners integrate the different sources of information into a coherent mental model	Instruction is adapted to the preferences of learners Learners can select information presented in different representational codes Learners can learn at their own pace	Learners do not recognize which fragment of information can close their knowledge gap Learners experience difficulties in making decisions about which content to access in what pace and order	Support for navigational and representational choices Levels of learner control combined with levels of prior knowledge
Observational learning	Learners observe experts performing a task or solving a problem while explicating and explaining the procedure and the rationale behind it	Learners rehearse the task mentally to build and refine their cognitive representation of the domain	Instruction can use realistic tasks to enhance motivation of the learners Learners are explicitly told about the decision processes and strategies, and explanations are given	Learners have no knowledge about which information in the perceptually rich environment is relevant Learners passively watch and listen to the expert instead of actively encoding the expert's information	Scaffolding whole-task practice Visual grouping Pacing Segmentation Type of role models

(continued)

Table 1 (continued)

	Description	Assumptions	Strengths	Risks	Instructional Variables
Self-explanation-based learning	A series of worked-out examples is presented and learners are encouraged to self-explain the solution steps of the examples	Learners use their cognitive resources to understand the solution steps presented Learners engage in effective self-explanation activities	Learners do not have to find the solution themselves and can use their cognitive resources to understand the solution steps presented Self-explaining shows which information is still missing	Learners experience difficulties in integrating and understanding the information presented Learners do not self-explain spontaneously	Saliency of subgoals in examples Support for integrating different representations used in the example Support for self-explanations
Inquiry learning	Learners induce characteristics of the domain from active experiences with the subject matter	Learners generate hypotheses, conduct experiments, and draw conclusions Learners are able to induce a general model from the concrete outcomes of the experiments performed	Thinking about the nature of the underlying model and conducting experiments to confirm or falsify their hypotheses leads to deep, conceptual understanding of the subject matter	Learners do not generate hypotheses Learners experience difficulties in performing suitable and sufficient experiments Learners draw invalid conclusions leading to incorrect knowledge	Types of assignments Model progression Support in making hypotheses Support in conducting experiments

frequently used the (static) graphical representations accompanying the problem situation and solution steps represented in text compared to learning from text only. However, frequent use of animations led to a decrease in performance and an increase in learning time compared to the text-only condition. For far transfer tasks, no differences were reported; conceptual knowledge was not measured in this study.

Observational Learning: Show Me

Observational learning involves observing others (experts) performing a task or solving a problem. This type of learning has traditionally been associated with behavior that is manifest and easily observable by the learner (Wetzel, Radtke, & Stern, 1994). The learner watches the expert and can see the processes at stake; it often involves learning physical activities, such as tailoring or playing tennis. Nowadays, however, observational learning is also applied in cognitive domains. The difference, though, is that the processes involved are not visible and thus must be made explicit by the expert (Collins, 1991). This means that the expert explicates decision processes and strategies underlying the problem-solving activities and also gives an explanation of why and when particular strategies are useful. This is thought to be effective for learning, as the emphasis in observational learning lies on the rationale behind the procedure as well as on the procedure itself (Collins, 1991; van Gog, Paas, & van Merriënboer, 2004). Learners can subsequently rehearse the task mentally and by so doing refine their initial representations (Bandura, 1976). Computer-based environments can present experts to learners as real experts but also as animated figures that show and explain expert behavior. One potential disadvantage of observational learning is that learners do not actively encode the information they receive from the expert, but watch the expert passively without actively trying to build or refine their cognitive representation of the domain. Several ways to overcome this problem have been proposed. "Scaffolding whole-task practice" gives students extensive support and guidance in the beginning of the learning process, which is faded as learners acquire more expertise (van Merriënboer, Kirschner, & Kester, 2003; van Merriënboer & Sweller, 2005). Other instructional variables address pacing (Schwan & Riemp, 2004), visual grouping (Rieber, 1993), and segmentation of the animation (Schwan & Garsoffky, 2004; Zacks & Tversky, 2001). In the domain of mathematics, Schunk and Hanson (1985) found that learners watching a videotape in which a peer solved subtraction problems scored higher on both near and far transfer tasks than learners watching a videotape in which a teacher applied the appropriate operations for several problems. Both groups outperformed learners in the control group, which received only a training program. In a later study, Schunk and Hanson (1989) found that learners watching themselves or peers solving fractions on videotape outperformed learners in several control conditions on a skill

test measuring near as well as far transfer. No conceptual knowledge measurements were included in either study.

Self-Explanation-Based Learning: Let Me Explain

Self-explanation-based learning leans on two ideas: example-based learning and self-explanation. Using worked-out examples in instruction involves introducing a principle, rule, or theorem; then providing a worked-out example consisting of a problem formulation, solution steps, and the solution; and finally supplying one or more problems to be solved. In example-based learning, the worked-out example gets a more central role in the instruction. Instead of only one example, a series of worked-out examples is presented. Learning from worked-out examples is assumed to be effective because the learners are freed from finding the solution on their own. Instead of looking for the right solution, they can use their cognitive resources to understand the solution steps (cf. Paas, Renkl, & Sweller, 2003). Self-explanation-based learning encourages learners to self-explain the solution steps of the worked-out examples to reach this understanding. Studies have shown that learners who explained the worked-out examples more actively to themselves learned more than learners who did not self-explain the examples (Chi et al., 1989; Renkl, 1997). Relevant instructional practices in self-explanation-based learning are making the subgoals in examples salient (Atkinson & Derry, 2000) and supporting the learners in integrating different representations used in the worked-out example (Tarmizi & Sweller, 1988). In the domain of probability theory, Atkinson, Renkl, and Merrill (2003) offered learners worked-out examples with the solution steps either faded or not and accompanied by self-explanation prompts or not. Data showed that a combination of fading solution steps and prompting learners to identify the underlying principle illustrated in the worked-out example resulted in good performance on both near transfer and far transfer tasks. Conceptual knowledge was not measured in this study.

Inquiry Learning: Let Me Investigate

Inquiry learning can be defined as a learning process in which learners induce characteristics of the domain from active experiences with the subject matter (de Jong, 2005; Swaak & de Jong, 1996). The learner can discover the concepts and variables in the underlying model that defines the properties of a particular domain by performing experiments and inferring information from the collected data. On the basis of an overview of a large set of studies, de Jong (2006b) introduced a set of specific learning processes that characterize inquiry learning. The main learning processes are: orientation, hypothesis generation, experimentation, and drawing conclusions. When the learning processes are correctly performed and the knowledge of the learner develops accordingly, inquiry learning can show promising results. However,

research has also shown that learners may encounter several difficulties in inquiry learning environments (de Jong & van Joolingen, 1998). For instance, learners may show “floundering behavior,” that is, doing experiments in an unsystematic way; they may find it hard to formulate hypotheses (Klayman & Ha, 1987; Njoo & de Jong, 1993); they may look only for evidence supporting their hypotheses instead of falsifying them (Quinn & Alessi, 1994); and they often draw conclusions that cannot be drawn from the evidence (Klahr & Dunbar, 1988). Therefore, learners need support in inquiry learning to make it an efficient and effective learning process (de Jong, 2005; de Jong & van Joolingen, 1998; Mayer, 2004). Examples of effective support tools include proposition tables facilitating comparison and discussion of multiple hypotheses (Gijlers & de Jong, in press), assignments presenting learners with short exercises, and model progression making complex domains more manageable (Swaak, van Joolingen, & de Jong, 1998). Studies on inquiry learning in the domain of mathematics are mainly embedded in educational settings. These studies show that learners who invented strategies before learning standard algorithms demonstrated better conceptual knowledge and were more successful in applying their knowledge to new situations than learners who immediately learned standard algorithms (e.g., Carpenter, Franke, Jacobs, Fennema, & Empson, 1998).

Hypotheses

This section describes the predictions about learning effects of each instructional approach. For each instructional approach, the learning processes that are assumed to be elicited by that approach are described as well as which knowledge type(s) are promoted by these learning processes. The classification scheme for study processes of Ferguson-Hessler and de Jong (1990) and the classes to code regulatory behavior of Azevedo, Cromley, and Seibert (2004) served as a starting point to identify learning processes (see Table 2 for an overview of learning processes assumed to be elicited by each instructional approach). The section ends with a comparison of the four instructional approaches in which predictions are given about learning effects related to specific types of knowledge.

Hypermedia Learning

In a hypermedia learning environment, the knowledge that a learner must acquire is available to the learner in different representational formats. In order to have access to the information, learners have to identify which knowledge they lack (Lawless & Brown, 1997). In order to do this, they have to relate the information they already selected to their prior knowledge and integrate the two into a coherent mental representation (Schnotz & Hei, 2009). Once they know which knowledge is missing, they can make intentional information selections (Barab, Bowdish, Young, & Owen, 1996), a process that requires self-regulating processes such as orienting, planning,

Table 2
Overview of Learning Processes Assumed to be Elicited by Each Instructional Approach

	Transformative Superficial		Transformative Elaboration				Regulative			
	Reading/ Watching and Filtering	Checking and Solution Steps	Developing and Testing Hypotheses	Relating and Integrating	Giving (Self- Explanations	Orienting	Planning	Monitoring	Reflecting	
HL	++	++	0	++	0	+	+	+	+	
OL	++	++	0	+	0	0	0	0	++	
EL	+	++	0	+	++	0	0	0	0	
IL	+	0	++	++	0	+	+	+	+	

Note. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

monitoring, and reflecting (Azevedo & Cromley, 2004). Once information has been selected, it has to be read and/or watched and filtered, and in case of problem solving, the solution steps must be checked. We expect that checking the solution steps of a problem in different representational formats leads to a flexible and firm knowledge base independent from representational format, which especially results in good procedural knowledge, that is, the ability to solve near transfer as well as far transfer problems.

Observational Learning

In an observational learning environment, a pedagogical agent presents the knowledge that must be acquired to the learner. In case of problem solving, this information concerns the solution steps and the rationale behind the problem-solving procedure (Collins, 1991; van Gog et al., 2004). The learner watches the animation, checks the solution steps, and reads or listens to why the problem is solved in that particular way. After each animation, the learner is supposed to reflect on the learning process by rehearsing the task mentally and integrating the newly received information into a mental representation (Wouters, Paas, & van Merriënboer, 2008). We expect that telling the learners explicitly about the underlying decision processes and strategies leads to good situational knowledge, that is, knowledge about typical situations and problem categories in a particular domain (de Jong & Ferguson-Hessler, 1996). Furthermore, we expect that watching an expert solving a problem and checking the solution steps simultaneously or rehearsing them afterwards leads to good procedural knowledge.

Self-Explanation-Based Learning

In a self-explanation-based learning environment, a combination of presenting subject matter and generating self-explanations is offered. The learner reads, watches, and filters the information provided and checks the solution steps. During this process, the learner is prompted to self-explain the solution steps. The prompts to self-explain make the learners think about underlying concepts and the rationale behind the procedure, make them generate inferences to fill in missing information or repair faulty knowledge, and make them relate and integrate the offered information together with prior knowledge into a mental representation (Roy & Chi, 2005). We expect that checking the solution steps, relating and integrating, and giving self-explanations leads to a complete knowledge base in which conceptual, intuitive, procedural (near as well as far transfer), and situational knowledge are represented.

Inquiry Learning

In an inquiry learning environment, the focus is on performing experiments in order to be able to induce knowledge from the learning environment.

The learner must orient on the environment and on prior knowledge related to the subject matter (de Jong & van Joolingen, 1998). Based on this information and supported by the guidance in the learning environment, the learner develops hypotheses, makes plans to test these hypotheses, and tests them by performing experiments (de Jong, 2006b). Comparing and relating the data from multiple experiments makes the underlying concepts and relations between variables salient to the learner. This new information must then be integrated into a mental representation and if after reflection it becomes clear that extra information is needed, new experiments can be performed (Klahr & Dunbar, 1988). We expect that relating data from different experiments makes clear which concepts and relations between variables are important, leading to a high quality of structured knowledge about concepts (i.e., conceptual knowledge and intuitive knowledge).

Comparison of Instructional Approaches

While assuming that each instructional approach will lead to learning effects, we particularly expect differences between approaches in the types of knowledge they promote. In the preceding sections, we described the characteristics of each instructional approach, the learning processes they are supposed to elicit, and predictions about which knowledge types these learning processes promote. The following predictions based on these sections result when comparing across instructional approaches: With regard to conceptual knowledge, we expect inquiry-based learners to score best, as learners perform experiments that show clearly the underlying concepts and relations between variables. The inquiry learning approach is assumed to be the only approach that focuses largely on conceptual rather than procedural knowledge. With regard to intuitive knowledge, we also expect inquiry-based learners to score best. Intuitive knowledge concerns a high-quality intuitive understanding of the subject matter (Swaak & de Jong, 2001a). It is hard to verbalize but by its intuitive character easier to access than conceptual knowledge. The action- and perception-driven elements in the inquiry learning environment are hypothesized to lead to high scores on intuitive knowledge. With regard to procedural knowledge, we expect learners in the hypermedia learning environment to score best. Learners can select and view the procedural information in different representational formats, leading to a flexible and firm knowledge base independent from representational format. This abstract knowledge about how to solve problems in the particular domain is useful for solving near as well as far transfer tasks. With respect to situational knowledge, we expect learners in the observational learning environment to score best. These learners are explicitly told why and when particular strategies are useful when solving problems. They learn to recognize typical problem situations and which problem-solving procedures to use in those cases. Finally, we expect learners in the self-explanation-based learning environment to score best overall.

Providing solution steps and explaining them in a systematic way leads to knowledge that can be used to categorize and solve problems. The prompts to self-explain make learners think about the underlying concepts and the rationale behind the procedures. The combination is expected to lead to a complete knowledge base consisting of conceptual, intuitive, procedural, and situational knowledge.

Method

The research described in this article consists of four studies, two performed in Germany and two performed in the Netherlands, and used a pretest/posttest quasi-experimental design. To ensure that the outcomes of the four studies could be compared, they all used the same domain (probability theory), the same procedure, the same introduction to the domain, the same learning material (i.e., the same set of concrete example situations), the same pre- and posttests (measuring conceptual, procedural, intuitive, and situational knowledge), and the same cognitive load measures. Materials for the Dutch studies were in Dutch and for the German studies in German. Details will be given in this section.

Participants

The data file on which this research is based contained the data of 624 participants (318 male, 303 female; 3 participants did not enter their sex). There were 365 participants in the German studies and 259 participants in the Dutch studies. All participants in the four studies were in Grades 10 or 11 of the highest level of secondary education. The curricula of both German and Dutch students were similar and prepared them for university. The mean age was 16.1 years ($SD = .9$). The number of participants working in the hypermedia learning environment, the observational learning environment, the self-explanation-based learning environment, and the inquiry learning environment was 196, 138, 169, and 121, respectively.

There were several cases where some data were not logged due to technical reasons and were thus not available for statistical analyses. This included data for learning times (3 participants), times to complete the posttest (24 participants), cognitive load measures during learning (12 participants for the distinctive cognitive load measures and an additional 13 participants for the overall cognitive load measure), and cognitive load measures during the posttest (20 participants). The data points involved were coded as missing data, which is reflected in fewer degrees of freedom in the descriptions of the corresponding statistical analyses.

The Domain

The domain of the experiment was combinatorics and elementary probability theory, which are treated in the curricula of our students. The

domain deals with situations involving the determination of the probability of randomly selecting a particular configuration of elements out of a set of elements (e.g., What is the probability that you correctly guess someone else's PIN code in one try?).

The probability of such a complex event depends on the total number of elements (i.e., the possible outcomes n), the number of elements that meet the selection criteria (i.e., the acceptable outcomes k), the number of selections to be made (i.e., the number of individual events), whether elements are replaced or not after selection, and whether the order of selection of elements matters. Combining the two variables of replacement and order gives four problem categories: (a) order important with replacement, (b) order not important with replacement, (c) order important without replacement, and (d) order not important without replacement. The probability can be calculated by either the individual events method or the formula-based method. In the individual events method the probability of each individual event is determined in succession and the probability of the complex event can be calculated by multiplying those of the individual events. In the formula-based method, a formula is used to determine the number of possible combinations A , after which the probability can be calculated.

Figure 1 shows a typical example of a problem in the domain of combinatorics and elementary probability theory. The figure makes clearer the difference between using the individual events method and using the formula-based method to solve this problem. In three of the learning environments, probabilities were calculated by the individual events method. The observation-based learning environment was the only environment that used both problem-solving procedures, as this learning environment mirrored expert problem-solving behavior and experts would use both procedures. The problem given in Figure 1 is taken from the posttest.

Learning Material

Research has shown that the use of examples assists in learning to solve mathematical problems (Catrambone, 1994; Cooper & Sweller, 1987; Zhu & Simon, 1987). All of our learning environments were therefore designed around a set of concrete example situations. Four examples were developed, one for each problem category. These examples concerned (a) guessing a PIN code (What is the probability that you guess the code correctly in one try?), (b) a marketing campaign in which collectible objects are given away together with a pack of muesli (What is the probability that you get the three objects you want most?), (c) predicting the first three finishers in a race (What is the probability that you correctly predict the persons in Positions 1, 2, and 3?), and (d) checking mobile phones on an assembly line for manufacturing errors (What is the probability that you correctly predict which phones will be checked?; see Figure 2 for an overview). Furthermore, a fifth description of a situation was developed (i.e., a 2-day mountain bike course in which helmets of different colors are handed out) that could be adapted to each of

Problem description:	
<p>In a music magazine you see an ad in the FOR SALE section offering a ticket for a spectacular concert by your favorite band. Unfortunately the last 2 digits of the telephone number where you can get information about the ticket are not readable anymore. You would really like to have the ticket and decide to choose the 2 digits randomly. What is the probability that you dial the correct digits on your first try?</p>	
Problem-Solving Procedures:	
<u>Individual events method:</u>	<u>Formula-based method:</u>
1. Problem category: order important with replacement	1. Problem category: order important with replacement
2. Probability first individual event (1/n): 1/10	2. Number of possible combinations ($A = n^k$): $10^2 = 100$
3. Probability second individual event (1/n): 1/10	3. Probability (1/A): 1/100
4. Probability complex event: $1/10 * 1/10 = 1/100$	

Figure 1. A typical example of a problem and the two corresponding problem-solving procedures.

		Order Important?	
		Yes	No
Replacement?	Yes	I. Guessing a pin code II. Predicting which color of helmet you get on the first day and which color of helmet you get on the second day	I. Collecting three different objects that are given away with each pack of muesli II. Predicting which color of helmets you get on the first and second day
	No	I. Predicting the persons in Positions 1, 2, and 3 II. Predicting which color of helmet you get and which your friend gets on the first day	I. Checking mobile phones on an assembly line for manufacturing errors II. Predicting which colors of helmets you and your friend get on the first day

Figure 2. The eight examples used in the experiment, each representing one of the four problem categories.

the four problem categories, resulting in another four examples. These last four examples were consequently identical on surface characteristics (i.e., handing out helmets for a mountain bike course) but different on underlying structure (i.e., the problem category; Quilici & Mayer, 1996).

Table 3
Overview of Pretest and Posttest Items

Knowledge Type	Knowledge Subtype	Description	Number and Type of Items
Pretest			
Conceptual	C1	Knowledge about effects of variables differing between problem categories	2 multiple choice items (4 alternatives)
	C2	Knowledge about effects of variables that can vary within problem categories	2 multiple choice items (4 alternatives)
Procedural	P2-easy	Knowledge about solving the most basic combinatorial problems	4 open-ended items
	P2-near	Knowledge about solving specific near transfer problems	4 open-ended items
Posttest			
Conceptual	C1	Knowledge about effects of variables differing between problem categories	4 multiple choice items (4 alternatives)
	C2	Knowledge about effects of variables that can vary within problem categories	4 multiple choice items (4 alternatives)
	C3	Understanding why probabilities are calculated the way they are	4 open-ended items
Intuitive	Cl _i	Knowledge about effects of variables differing between problem categories	13 multiple choice items (2 alternatives)
Procedural	P1	Knowledge about calculating probabilities	2 open-ended items
	P2-near	Knowledge about solving specific near transfer problems about probability	8 open-ended items
	P2-far	Knowledge about solving specific far transfer problems about probability	4 open-ended items
Situational	Sit ₁	Knowledge about identifying relevant information in a problem	1 open-ended item and 4 multiple choice items (4 alternatives)

Questionnaire and Tests

Participant characteristics questionnaire. A participant characteristics questionnaire was administered consisting of questions concerning personal data (age, gender, etc.); questions concerning prior experience with probability theory, computers, and computer-based learning environments; and questions concerning self-ratings of verbal and spatial abilities and multimedia preferences.

Pretest and posttest. A pretest and posttest were administered to measure the learners' knowledge about probability theory. Because different instructional approaches can be more or less sensitive to different kinds of knowledge, we included a variety of different knowledge measures in the pretest and the posttest. Table 3 gives an overview of the pretest and posttest items. The pretest consisted of 12 items measuring conceptual and procedural knowledge. The 4 conceptual items measured knowledge about the effect of variables that differ between problem categories, such as order and replacement (C1), and knowledge about the effect of variables that can vary within problem categories, such as the number of possible outcomes, the number of acceptable outcomes, and the number of selections to be made (C2). The 8 procedural items measured basic everyday life knowledge about very simple probability situations (P2-easy) and the ability to apply this knowledge to less simple problem-solving situations (P2-near). More complex items that were used in the posttest (P2-far) were not used in the pretest. All items in the pretest except those related to everyday knowledge (P2-easy) had twin items in the posttest so that the knowledge gain of the learners could be determined. These twin items shared the same structure but differed at a superficial level, such as order of alternatives and names. The 4 basic everyday life items (P2-easy) on the pretest were used to assess the commonsense prior knowledge of the learners that is a prerequisite for learning probability theory.

The posttest consisted of 44 items measuring conceptual, intuitive, procedural, and situational knowledge (see Table 4 for typical examples of each item type). The 12 items measuring conceptual knowledge included 4 items about the effect of variables differing between problem categories (C1), 4 items about the effect of variables that can vary within problem categories (C2), and 4 items about why problems have to be solved in a particular way (C3). The 13 items measuring intuitive knowledge concerned the effect of variables differing between problem categories (C1i). Intuitive knowledge refers to intuitive or implicit ideas of the subject matter that anticipate the outcome of an argument and that are hard to verbalize. There were three differences between measuring intuitive knowledge and C1 conceptual knowledge: (a) The problem situation in the intuitive items was the same for each item and was presented prior to the items instead of given within the items, (b) the intuitive items offered two alternatives instead of four, and most importantly (c) learners were asked to answer the intuitive items as quickly as possible, as intuitive knowledge is characterized by a quick perception of meaningful situations (Swaak & de Jong, 1996). The 14 items measuring procedural knowledge included 2 items in which learners had to describe how to calculate probabilities (P1), 8 items in which they had to solve a specific near transfer problem (P2-near), and 4 items in which they had to solve a specific far transfer problem (P2-far). Finally, the 5 items measuring situational knowledge concerned items in which relevant information had to be filtered from a given problem situation (Sit1).

Table 4
Typical Examples of Each Item Type on the Posttest

Item Type	Typical Example
C1	<p>You have a deck of cards from which you select 4 cards. You want to get an ace, king, queen, and jack in this specific order. Does it matter whether you put back the selected cards before each new selection or not?</p> <ol style="list-style-type: none"> a. Yes, your chances increase when you put back the selected cards b. Yes, your chances decrease when you put back the selected cards c. No, your chances remain the same whether you put back the selected cards or not d. This depends on the number of jokers in the deck of cards
C2	<p>You visit the horse races. It's your first time and you have no idea which horses are good. You predict which horse will win and you make a bet on that. The race starts and there is one horse that wins with a large lead. All the other horses are still in the race and the differences are minimal. You can still make a bet on who will come in second. Do you have a greater chance of correctly predicting the second-place finisher compared to the chance you had to predict the winner?</p> <ol style="list-style-type: none"> a. Yes, the chance increases with my second bet b. No, the chance is the same for my second bet as it was for my first bet c. No, my chance decreases with my second bet d. There is no systematic relation between my first and my second bet
C3	<p>All 26 letters of the alphabet are collected in a vase. You must select 2 letters from the vase while blindfolded. After you select the first letter, it is set aside. The number of letter pairs you can select depends on whether the order in which you select them is important or not. Please explain this relation.</p>
Cl1	<p>The chance that your prediction comes true is greater with selections:^a</p> <ol style="list-style-type: none"> a. without replacement; order not important b. without replacement; order important
P1	<p>Describe in your own words how the probability of an event can be calculated in terms of acceptable outcomes and possible outcomes.</p>
P2-near	<p>Seven students take a Latin course. Lots are drawn to determine which 2 students should read their translation aloud. Stupidly, you have copied your translation from your friend. What is the probability that the two of you are chosen?</p>
P2-far	<p>In order to determine the final mark for a subject your teacher uses two complementary methods. First, you must perform 2 practice tasks, followed by a test consisting of 8 multiple choice questions. Two of 5 possible practice tasks (named A, B, C, D, and E) are randomly assigned to you. You know you did practice tasks D and E a month ago. For the multiple choice questions your teacher uses a large pool of 100 different multiple choice questions from which he randomly selects 8 questions for you. You have also already taken a test with 8 questions from this pool. What is the probability that you are assigned practice task D and E as well as the 8 questions you have had before?</p>
Sit	<p>You had a party and bought balloons in different colors. Now the party is over, you can pop the balloons. You will do this blindfolded and you predict that you will pop a red one first, then a yellow one, and finally a blue one. What is the characterization of this problem?</p> <ol style="list-style-type: none"> a. order important; replacement b. order important; no replacement c. order not important; replacement d. order not important; no replacement

^aPrior to the Cl1 items, an instruction was given to answer the items as quickly as possible and a general problem situation was given that was the same for each item.

The pretest and posttest were validated in two pilot studies. We started off with a total set of 69 items. The aim of the first pilot study was to check whether the items were suitable for the target group of students in Grades 10 and 11 of the highest level of secondary education. After each pretest and posttest item, three evaluation questions were presented to be answered on a 5-point scale. The first question concerned to what extent the learners understood the item, the second question to what extent they liked the content of the item, and the third question determined to what extent they found the item difficult or easy to answer. Higher scores indicated more understanding, more appreciation, and less perceived difficulty. Participants in this study were 12 students in Grade 10 of the highest level of secondary education (5 male, 7 female). The mean age was 15.0 years ($SD = .5$). The reliabilities of the pretest and posttest, as measured with Cronbach's α , were $\alpha = .37$ and $\alpha = .73$, respectively. The mean scores for understanding, appreciation, and perceived difficulty were 3.75 ($SD = .65$), 3.31 ($SD = .69$), and 3.22 ($SD = .52$). All means were slightly above the neutral position on the scale, indicating that students did not experience major problems with understanding, appreciation, and difficulty of the items. Analyses for the separate types of knowledge revealed a similar picture, although the far transfer items were experienced as difficult ($M = 2.44$, $SD = .54$) and the situational items were seen as less difficult ($M = 3.90$, $SD = .70$).

The aim of the second study was to assess the quality of the items in terms of p values, item-test correlations, and distribution of answers. Participants in this study were 50 students in Grade 12 of the highest level of secondary education (20 male, 30 female). The mean age was 16.4 years ($SD = .53$). All participants had received an introduction to probability theory and finished courses in probability theory as part of their regular curriculum in Grade 11, so that this group of students would match the target group for level of knowledge after completing the learning environment. The reliabilities of the pretest and posttest, as measured with Cronbach's α , were $\alpha = .41$ and $\alpha = .67$, respectively. On average, learners needed 51.9 minutes ($SD = 13.1$; range = 26.6–86.4 minutes) to complete all items. As this was considered too long, we decided to delete several items from the pretest and posttest in order to decrease the time needed to finish the tests. To reach an optimal trade-off between time on task and a valid and reliable posttest, a minimum number of items for each subtype of knowledge was defined. Criteria for deletion were based on statistical grounds, such as low p values, low or negative item-test correlations, and skewed distribution of responses over the different alternatives in the multiple choice questions. Almost all learners correctly answered 4 items in the pretest. Therefore, these items were excluded, resulting in a final pretest of 12 items. No items in the pretest were reformulated. In addition, 9 items in the posttest were deleted based on statistical grounds and 6 items were reformulated. This resulted in the final posttest of 44 items. The reliabilities of the pretest and posttest as used in the present study, measured with Cronbach's α , were $\alpha = .64$ and $\alpha = .82$, respectively.

Table 5

Overview of Cognitive Load Items Used in the Learning Environments

Type of Cognitive Load	Corresponding Item
Intrinsic load	How easy or difficult do you consider probability theory at this moment?
Extraneous load: navigation	How easy or difficult is it for you to work with the learning environment?
Extraneous load: design of the learning task	How easy or difficult is it for you to distinguish important and unimportant information in the learning environment?
Extraneous load: accessibility of information	How easy or difficult is it for you to collect all the information that you need in the learning environment?
Germane load	How easy or difficult was it to understand the simulation? ^a
Overall load	Indicate on the scale the amount of effort you had to invest to follow the last simulation. ^a

^aThis item was taken from the inquiry learning environment. The other environments used similar wordings depending on their particular instructional approach.

Cognitive load measures. Cognitive load was measured during the learning phase as well as during the posttest. During the learning phase, cognitive load was measured with 6 items assessing intrinsic, extraneous, germane, and overall load (see Table 5 for an overview). This set of cognitive load items was an adapted and extended version of the SOS scale (Swaak & de Jong, 2001b). Intrinsic load concerned the amount of mental effort with regard to the perceived difficulty of the domain in general. Extraneous load was measured by 3 items about the navigation, the design of the learning task, and the accessibility of information. Germane load concerned the understanding of the knowledge within the learning environment. Finally, the overall load concerned the amount of mental effort invested. The learners rated the items on a 9-point Likert scale, with higher scores indicating greater load. A set of cognitive load items was presented each time the learner went through one of the four problem categories. The items within the sets were presented in a different order each time to prevent learners from answering automatically.

During the posttest, cognitive load was measured with 1 item asking learners to indicate on a 9-point Likert scale the amount of effort they had to invest to solve the last problem. The item used was the same item used by Paas (1992) in his experiment. As it would be too much to present the cognitive load item after every posttest item, it was decided to administer the item 18 times, that is, after all open-ended items (except for 2 conceptual items) and once after completion of all intuitive items.

Procedure

Participants first completed the participant characteristics questionnaire. Then they were given a short introduction to the session and some technical

details regarding the use of the learning environment. After that, the pretest was administered. Questions were presented screen by screen without the possibility of skipping questions or returning to previous questions. An on-screen calculator was provided when learners had to calculate an answer. After completing the pretest, participants received an introduction to the domain of probability including the basic notion of random experiments and the general rationale behind calculating the probability of outcomes. The introduction was followed by the assigned learning environment, in which participants worked through the four problem categories. While working in the learning environment, the participants estimated the cognitive load they were experiencing. After the learning phase, the posttest, again including cognitive load measures, was administered. Aside from the learning environment, all procedural elements were identical across the four experimental studies.

Learning Environments

Four computer-based learning environments giving an introduction to probability theory were used. Each learning environment used a specific instructional approach: (a) hypermedia learning, (b) observational learning, (c) self-explanation-based learning, and (d) inquiry learning. Within each approach, specific variations were used. The variations concerned instructional variables such as representational format and amount of learner control. Each variation within an instructional approach was assumed to elicit the learning processes related to that approach. That is, the variations within an instructional approach were specific implementations and together they represented a particular approach. The learning environments with their specific variations are described next.

In the hypermedia learning environment, the solution steps for calculating the probability of a complex event were given by explaining each individual event in succession. The solution steps were given as a formula; or as a formula complemented with a textual representation, an animation, or an audio file; or a combination of these. Learners either received high learner control or low learner control. In the high learner control version, learners could either select or ignore the information given. This resulted in seven variations of this instructional approach. See Gerjets, Scheiter, Opfermann, Hesse, and Eysink (2009) for more detailed information on the learning environment and instructional variations.

The observational learning environment used an animated dolphin as a pedagogical agent. This dolphin moved within an animation representing the problem situation. The dolphin described the solution steps (either in text or in audio) to solve the problem and gave information about the underlying rationale for the steps. As the learning environment represents the way experts would solve the problems, both the individual event method and the formula-based approach of probability theory were used. The animations in the learning environment were either segmented or continuous and learner

paced or system paced, resulting in eight different versions of this instructional approach. See Wouters, Paas, and van Merriënboer (2006) for more information on the learning environment and instructional variations.

The self-explanation-based learning environment used three consecutive steps to explain the solution procedure for each problem situation. First, the acceptable outcomes were addressed, followed by why one has to multiply, and then finally by the possible outcomes. The solution steps were represented as a formula, as a tree diagram, or as a combination of a formula and a tree diagram. When combined, the learners either received no help in integrating the formula and tree diagram or they received an integration aid. The latter was accomplished by having the corresponding information from the two representations flashing simultaneously in the same color. Finally, learners either received scaffolded self-explanation prompts or no prompts. Combining these variables resulted in eight different instructional variations. For more detailed information, see Berthold and Renkl (2009) and Berthold, Eysink, and Renkl (2009).

In the inquiry learning environment, learners were presented with a simulation for each problem situation. The learners could vary different variables (e.g., the number of possible outcomes and the number of acceptable outcomes) and see the resulting effect on the probability. Learners were guided in their discovery processes by assignments. The solution steps were given either as a formula, a tree diagram, a textual description, a combined formula and tree diagram, or a combined formula and textual description resulting in five instructional variations. More details about the inquiry learning environment can be obtained from Kolloffel, Eysink, de Jong, and Wilhelm (in press).

Data Analyses

The data from the four studies conducted in Germany and the Netherlands were combined for the analyses presented in this article. Our aim was to compare instructional approaches independent of specific variations within these approaches. Therefore, we used all data gathered in the four studies; that is, data from all instructional variations within an instructional approach were included. In the data analyses, the data from the variations within each instructional approach were put together as representing that specific approach. For instance, no distinction was made between the eight instructional variations of the observational learning environment; they were all combined, resulting in one overall condition representing the observational learning instructional approach. This resulted in four experimental conditions corresponding to the four instructional approaches. The rationale for this is that statements about instructional approaches go beyond specific implementations of the approaches.

Data consisted of knowledge scores, times, and cognitive load scores. Learners received 1 point for each correct answer, resulting in a maximum score of 12 for the pretest and 44 for the posttest. Times were logged, so that

Table 6

Effectiveness and Efficiency Indicators Used to Determine Differences Between Instructional Approaches

Effectiveness of instructional approach	
Effectiveness I	Posttest score
Effectiveness II	$\frac{\text{Posttest score}}{\text{Cognitive load during posttest}}$
Efficiency of learning with an instructional approach	
Efficiency I	$\frac{\text{Effectiveness I}}{\text{Learning time}}$
Efficiency II	$\frac{\text{Effectiveness II}}{\text{Learning time}}$

both the time learners spent in the learning environment and the time they took to complete the posttest could be computed. Furthermore, cognitive load scores were logged during learning and during the posttest.

Various indicators were used to assess differences between instructional approaches (see Table 6). The indicators represent the effectiveness of the instructional approaches as well as the efficiency of the learning process.

Effectiveness of instructional approaches concerns the effect of an instructional approach on achievement. The first effectiveness indicator (Effectiveness I) assesses achievement by test performance examined at the levels of the total posttest score and the scores for each subtype of knowledge represented in the posttest (see Table 3). The second effectiveness indicator (Effectiveness II) assumes that learners with better organized cognitive schemata need less mental effort to solve a problem (Kalyuga & Sweller, 2005; Yeo & Neal, 2004). In other words, if a learner in one instructional approach needs less mental effort to attain equal or higher test performance, then the learner's cognitive schema is better organized, which in turn means that the instructional approach that was used for learning was more effective. This is similar to what Paas and van Merriënboer (1993) and van Gog and Paas (2008) called *instructional efficiency*. In their measure, they take the difference between the standardized z scores of performance and the standardized z scores of mental effort involved in that performance and they divide this by $\sqrt{2}$. Besides that we do not feel that the term *instructional efficiency* gets at the feel of this measure, we propose a different and more intuitive and simple indicator for the same construct. We define the second effectiveness indicator as the ratio of the posttest score to the cognitive load measured during solving the posttest. We deliberately do not use standardized z scores, as we want a measure indicating the effectiveness of a certain instructional approach independent of other instructional approaches.

Efficiency of learning with an instructional approach concerns the amount of time required to reach a certain level of achievement (see also e.g., Cates et al., 2003; Schmidgall & Joseph, 2007). As with the effectiveness indicators, the first efficiency indicator (Efficiency I) assesses achievement by test performance, whereas the second efficiency indicator (Efficiency II) assesses achievement by the posttest score in relation to the mental effort it takes to reach this score. Time in this instance concerns time learners spent in the learning environment. We define both efficiency indicators I and II as the ratio of the corresponding effectiveness indicator I or II to learning time. We do not use log functions of learning time (see e.g., Pietrzak, Cohen, & Snyder, 2007), as it is assumed (and confirmed by the data) that learners do not reach the maximum score that can be attained on the posttest. Furthermore, we do not include cognitive load during learning in the efficiency measurements as the exact rationale behind this idea is not well described in the literature, and we could not find an unambiguous answer to the question of which type of load should be used for such a measurement.

Results

Covariates

Table 7 gives the mean pretest scores for the learners in each instructional approach. The relatively high scores can be explained by the fact that the pretest included easy items measuring basic knowledge about chances acquired during everyday life (P2-easy). A one-way ANOVA showed significant differences between instructional approaches on the pretest scores, $F(3, 620) = 55.37, p < .001$, partial $\eta^2 = .21$. Therefore, pretest score was taken into account as a covariate in further analyses.

Effectiveness I

Table 8 gives the posttest means for each instructional approach overall and for the individual parts reflecting the subtypes of knowledge. All means are corrected for pretest scores. This section describes the effects of each instructional approach for the different subtypes of knowledge. For the posttest in total, a one-way ANCOVA was performed. For the subtypes of knowledge, a MANCOVA was performed.

Posttest total. Results show a main effect of instructional approach for total posttest score, $F(3, 623) = 28.94, p < .001$, partial $\eta^2 = .12$. Pairwise comparisons using the Bonferroni procedure show that learners in the self-explanation-based learning environment scored significantly higher than those in all other conditions (all $ps \leq .001$), and learners in the inquiry learning environment scored significantly higher than those in the hypermedia learning environment ($p < .01$) and the observational learning environment ($p < .05$).

Table 7
Mean Pretest Scores and Standard Deviations

	Instructional Approach							
	HL		OL		EL		IL	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pretest score	6.74	1.98	6.36	2.10	5.07	2.19	8.12	1.73

Note. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

Table 8
Means Corrected for Pretest Scores and Standard Errors for the Subtypes of Knowledge in the Posttest and the Posttest in Total for Each Instructional Approach

Subtype of Knowledge	Instructional Approach							
	HL		OL		EL		IL	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Total	23.66	.38	23.67	.45	28.47	.43	25.72	.51
C1: effects between categories	2.71	.07	2.45	.08	3.09	.08	2.77	.09
C2: effects within categories	2.84	.06	2.74	.07	3.15	.07	2.98	.08
C3: understanding why	1.45	.07	1.44	.08	2.10	.08	1.69	.09
C1i: intuitive knowledge	8.98	.19	9.05	.22	9.96	.21	9.35	.25
P1: knowing how	0.76	.04	0.39	.05	0.94	.05	0.41	.06
P2-near: near transfer	3.35	.12	3.72	.14	4.69	.14	3.94	.17
P2-far: far transfer	0.50	.06	0.81	.07	0.93	.07	1.28	.08
Sit: situational knowledge	3.07	.10	3.08	.12	3.60	.11	3.30	.13

Note. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

Conceptual knowledge. A main effect was found for conceptual knowledge of effects of variables differing between problem categories (C1), $F(3, 623) = 11.12, p < .001$, partial $\eta^2 = .05$. Pairwise comparisons using the Bonferroni procedure showed that learners in the self-explanation-based learning environment scored higher than those in the hypermedia learning environment ($p < .01$) and in the observation-based learning environment ($p < .001$). No significant differences were found between learners in the self-explanation-based learning environment and those in the inquiry learning environment on the one hand and learners in the inquiry learning environment and those in the hypermedia and observational learning environment on the other hand.

A main effect was also found for conceptual knowledge of effects of variables that can vary within problem categories (C2), $F(3, 623) = 7.08, p < .001$, partial $\eta^2 = .03$. Again, learners in the self-explanation-based learning environment scored significantly higher than those in the hypermedia learning environment ($p < .01$) and in the observation-based learning environment ($p < .001$). Again, no significant differences were found between self-explanation-based learning and inquiry learning on the one hand and between inquiry learning and hypermedia and observational learning on the other hand.

A third main effect was found for conceptual knowledge of why to calculate the probability in a particular way (C3), $F(3, 623) = 17.08, p < .001$, partial $\eta^2 = .08$. Pairwise comparisons showed that learners in the self-explanation-based learning environment scored significantly higher than all three other learning environments ($p < .01$ for inquiry learning and $p < .001$ for hypermedia learning and observation-based learning).

Intuitive knowledge. A main effect was found between instructional approaches for intuitive knowledge of effects of variables differing between problem categories (C1i), $F(3, 623) = 4.72, p < .01$, partial $\eta^2 = .02$. Pairwise comparisons using the Bonferroni procedure showed that learners in the self-explanation-based learning environment outperformed learners in the observation-based learning environment ($p < .05$) and those in the hypermedia learning environment ($p < .01$). No significant differences were found between learners in the inquiry learning environment and those in the other learning environments.

Procedural knowledge. A main effect was found for procedural knowledge of how to calculate probabilities (P1), $F(3, 623) = 28.41, p < .001$, partial $\eta^2 = .12$. Pairwise comparisons using the Bonferroni procedure showed that this effect was caused by the fact that learners in the self-explanation-based learning environment scored significantly higher than learners in all other three conditions ($p < .05$ for hypermedia learning and $p < .001$ for inquiry and observational learning), and learners in the hypermedia learning environment scored significantly higher than learners in the inquiry and observational learning environment (both $ps < .001$).

Procedural knowledge of near transfer items (P2-near) also differed between instructional approaches, $F(3, 623) = 17.83, p < .001$, partial $\eta^2 = .08$. Pairwise comparisons showed that learners in the self-explanation-based learning environment outperformed those in the other three instructional approaches ($p < .01$ for inquiry learning and $p < .001$ for observational learning and hypermedia learning). Furthermore, learners in the inquiry learning environment scored significantly higher than those in the hypermedia learning environment ($p < .05$).

A main effect was found as well for the third type of procedural knowledge, procedural knowledge of far transfer items (P2-far), $F(3, 623) = 21.19, p < .001$, partial $\eta^2 = .09$. This time, learners in the inquiry learning environment

outperformed learners in the other three conditions ($p < .05$ for self-explanation-based learning and $p < .001$ for observational and hypermedia learning). Furthermore, both self-explanation-based and observational learners scored significantly higher than learners in the hypermedia learning environment ($p < .001$ and $p < .01$, respectively).

Situational knowledge. A main effect was found for situational knowledge of identifying relevant information from a problem (Sit), $F(3, 623) = 5.29$, $p \leq .001$, partial $\eta^2 = .03$. Pairwise comparisons using the Bonferroni procedure show that this effect can be attributed to learners in the self-explanation-based learning environment scoring higher than those in the observation-based and the hypermedia learning environments (both $ps < .01$). Learners in the inquiry learning environment did not differ significantly from learners in the self-explanation-based learning environment or from those in the observational and hypermedia learning environment.

Effectiveness II

The second effectiveness indicator concerns the mental effort it takes to attain a given level of achievement in the posttest (van Gog & Paas, 2008). Table 9 gives the mean cognitive load scores during the posttest. Results show a main effect of instructional approach for cognitive load during the posttest, $F(3, 603) = 9.63$, $p < .001$, partial $\eta^2 = .05$. Pairwise comparisons using the Bonferroni procedure show that learners in the observational learning environment experienced significantly higher loads than those in the inquiry learning environment ($p < .05$) and in the self-explanation-based learning environment ($p < .001$). Furthermore, learners in the hypermedia learning environment experienced significantly higher loads than those in the self-explanation-based learning environment ($p \leq .001$).

We defined Effectiveness II as the ratio of the posttest score to the cognitive load measured during the posttest. Table 10 gives the mean scores for each instructional approach on Effectiveness II. The means are corrected for pretest scores. Results show a main effect of instructional approach for Effectiveness II, $F(3, 603) = 24.63$, $p < .001$, partial $\eta^2 = .11$. Pairwise comparisons using the Bonferroni procedure show that learners in the self-explanation-based learning environment scored significantly higher than learners in all other conditions (all $ps < .001$).

The second effectiveness indicator is intended to measure the effort a learner needs to reach a certain level of achievement. In our case, effort is defined as cognitive load while completing the posttest. However, one could also argue that effort can be seen as the mental effort in combination with the time spent to achieve the given result (cf. van Gog et al., 2005). The rationale behind this argument is that individuals with more expertise will need less mental effort as well as less time to attain equal or even higher levels of performance. Table 9 gives the mean times learners in the four instructional approaches needed to complete the posttest. Results show a

Table 9
Means and Standard Deviations for Cognitive Load and Times in Absolute Minutes for Each Instructional Approach

	Instructional Approach							
	HL		OL		EL		IL	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cognitive load while completing posttest	4.78	1.53	5.07	1.51	4.15	1.53	4.52	1.63
Time needed to complete posttest	40.78	11.95	33.14	9.86	55.54	17.47	32.58	9.25
Time used for learning	18.27	6.48	29.93	6.90	55.98	18.93	33.33	9.43

Note. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

Table 10
Means Corrected for Pretest Scores and Standard Errors for the Effectiveness and Efficiency Measurements for Each Instructional Approach

	Instructional Approach							
	HL		OL		EL		IL	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Effectiveness								
Effectiveness I	23.66	.38	23.67	.45	28.47	.43	25.72	.51
Effectiveness II	5.71	.25	5.57	.30	8.61	.28	6.20	.35
Effectiveness III	0.15	.01	0.18	.01	0.17	.01	0.21	.01
Efficiency								
Efficiency I	1.47	.03	0.82	.04	0.64	.04	0.81	.04
Efficiency II	0.37	.02	0.19	.02	0.20	.02	0.19	.02

Note. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

main effect of instructional approach for times, $F(3, 599) = 100.44, p < .001$, partial $\eta^2 = .34$. Pairwise comparisons using the Bonferroni procedure show that learners in the self-explanation-based learning environment needed more time than those in all three other environments (all $ps < .001$). Furthermore, learners in the hypermedia learning environment needed more time to complete the posttest than those in the observational learning environment and the inquiry learning environment (both $ps < .001$).

When time is included in the second effectiveness indicator, this measurement is defined as the ratio of posttest performance to the product of cognitive load and time to complete the posttest. Table 10 gives the means of this third effectiveness indicator (Effectiveness III). Results show a main effect of instructional approach for Effectiveness III, $F(3, 597) = 7.63$,

$p < .001$, partial $\eta^2 = .04$. Pairwise comparisons using the Bonferroni procedure show that learners in the inquiry learning environment outperformed those in the self-explanation-based and the hypermedia learning environments (both $ps < .05$). Learners in the observational learning environment scored higher on Effectiveness III than those in the hypermedia learning environment ($p < .05$).

Efficiency I

Efficiency of learning with an instructional approach concerns the amount of time necessary to reach a certain level of achievement. Table 9 gives the mean times learners in the four instructional approaches spent on learning. Results show a main effect of instructional approach for learning time, $F(3, 620) = 318.73$, $p < .001$, partial $\eta^2 = .61$. Pairwise comparisons using the Bonferroni procedure show that learners in the self-explanation-based learning environment needed significantly more time for learning than those in all other conditions ($p < .001$ for all conditions). Furthermore, learners in the inquiry learning environment and in the observational learning environment took significantly more time for learning than those in the hypermedia learning condition (both $ps < .001$).

The first efficiency indicator is defined as the ratio of test performance to learning time. Table 10 gives the means for this indicator for each instructional approach. Results show a main effect of instructional approach for Efficiency I, $F(3, 620) = 113.45$, $p < .001$, partial $\eta^2 = .36$. Pairwise comparisons using the Bonferroni procedure show that learners in the hypermedia learning environment were significantly more efficient than learners in all other conditions ($p < .001$ for all conditions). Learners in the observational learning environment and in the inquiry learning environment were more efficient than those in the self-explanation-based learning environment ($p < .01$ for observational learning and $p < .05$ for inquiry learning).

Efficiency II

The second efficiency indicator is defined as the ratio of Effectiveness II to learning time. Results show a main effect of instructional approach for Efficiency II, $F(3, 601) = 30.00$, $p < .001$, partial $\eta^2 = .13$. Pairwise comparisons using the Bonferroni procedure show that learners in the hypermedia learning environment were significantly more efficient than learners in all other conditions ($p < .001$ for all conditions).

Cognitive Load During Learning

When learning in a complex multimedia learning environment, the issue of cognitive load is an important one. In order to check whether cognitive load had an intermediate effect on learning outcomes, cognitive load during learning is described. Learners' responses to the cognitive load questions during learning were averaged for each type of cognitive load for each

Table 11
Mean Cognitive Load Measures and Standard Deviations
During Learning for Each Instructional Approach

Cognitive Load	Instructional Approach							
	HL		OL		EL		IL	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Intrinsic	3.02	1.50	3.25	1.81	3.70	1.93	3.71	1.43
Extraneous	2.66	1.21	2.44	1.41	3.58	1.56	3.66	1.20
Germane	2.31	1.25	2.13	1.41	3.18	1.84	3.61	1.35
Overall	3.53	2.09	2.22	1.48	5.04	2.36	3.56	1.35

Note. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

instructional approach. The three types of extraneous cognitive load (navigation, design, and accessibility) were taken together in these analyses. The result is given in Table 11.

A main effect was found for intrinsic cognitive load, $F(3, 611) = 6.72$, $p < .001$, partial $\eta^2 = .03$. Bonferroni post hoc analyses showed that learners in the hypermedia learning environment reported a significantly lower intrinsic cognitive load than those in the inquiry learning environment ($p < .01$) and the self-explanation-based learning environment ($p \leq .001$).

A main effect was found as well for extraneous load, $F(3, 611) = 31.03$, $p < .001$, partial $\eta^2 = .13$. Bonferroni post hoc analyses showed that learners in the hypermedia learning environment as well as in the observational learning environment reported significantly lower extraneous load than those in the self-explanation-based and inquiry learning environments ($p < .001$ for all comparisons).

Results also showed a main effect for germane cognitive load, $F(3, 611) = 32.02$, $p < .001$, partial $\eta^2 = .14$. Bonferroni post hoc analyses showed lower germane load for the hypermedia learners and the observational learners compared to the inquiry and the self-explanation-based learners (all $ps < .001$).

Finally, main effects were also found for overall cognitive load, $F(3, 598) = 50.92$, $p < .001$, partial $\eta^2 = .20$. Learners in the self-explanation-based learning environment reported significantly higher overall load than those in all other learning environments (all $ps < .001$). Furthermore, learners in the observational learning environment reported lower overall load than those in all other learning environments (all $ps < .001$).

Discussion

The purpose of this research was to compare different instructional approaches. Four instructional approaches were selected for comparison:

Table 12
Overview of Effects for Each Subtype of Knowledge

Subtype of Knowledge	Effects			
Total	EL >	(IL >	(OL =	HL))
C1: effects between categories	EL ≥	(IL =	(HL =	OL))
C2: effects within categories	EL ≥	(IL =	(HL =	OL))
C3: understanding why	EL >	(IL =	HL =	OL)
C1i: intuitive knowledge	EL ≥	(IL =	(OL =	HL))
P1: knowing how	EL >	(HL >	(IL =	OL))
P2-near: near transfer	EL >	(IL ≥	(OL =	HL))
P2-far: far transfer	IL >	((EL =	OL) >	HL)
Sit: situational knowledge	EL ≥	(IL =	(OL =	HL))

Note. The symbol > represents a significant difference; the symbol ≥ represents no significant difference with the first following but a significant difference with the next following; the symbol = represents no significant difference. HL = hypermedia learning; OL = observational learning; EL = self-explanation-based learning; IL = inquiry learning.

(a) hypermedia learning, (b) observational learning, (c) self-explanation-based learning, and (d) inquiry learning. All four approaches are based on the idea that learners should engage in cognitive processes that are the basis for meaningful learning, such as selecting relevant information, organizing it into a coherent mental representation, and integrating it with other information and prior knowledge (Mayer, 2003). As a consequence, each instructional approach claims to lead to deep understanding and good learning results. In the current study, these approaches were compared for effectiveness as well as for efficiency. The instructional approaches were implemented in computer-based learning environments in such a way that the comparisons were fair. Each instructional approach followed the latest insights in instructional design and no caricature approaches were used, which led to realistic learning environments. The development and use of common elements (i.e., the use of the same learning material, the same procedure, the same tests to measure knowledge, the same measurements for cognitive load, participants from the same population, and the lack of the influence of teachers) in such a large-scale quasi-experimental setting made the research described in this article innovative and unique.

Assessing Instructional Approaches on Effectiveness

In order to assess the effectiveness of instructional approaches, approaches were compared on two effectiveness measures. Table 12 summarizes the effects that were found. Overall, it can be concluded that self-explanation-based learning is the most effective of the four instructional approaches that were examined, followed by inquiry learning. Hypermedia learning and observational learning score the lowest. This means that learners in the self-explanation-based learning environment acquire most knowledge and that the cognitive

schemata of these learners are organized most effectively. This matches our prediction that the combination of providing and explaining solution steps with prompts to self-explain leads to a complete knowledge base consisting of conceptual, intuitive, procedural, and situational knowledge. However, we should note as a criticism that the observed effect sizes were small to moderate.

In order to determine whether effects on different types of knowledge could be identified, approaches were compared on (the subtypes of) conceptual knowledge, intuitive knowledge, procedural knowledge, and situational knowledge. The overall pattern of self-explanation-based learning outperforming the other three approaches and inquiry learning scoring higher than both hypermedia and observational learning is seen for most subtypes of knowledge.

Results show that learners in the self-explanation-based learning environment ended up with more conceptual knowledge than learners in the hypermedia learning environment and the observational learning environment. Compared to the hypermedia and observational learners, self-explanation-based learners better understood the effect of variables differing within and between problem categories and they better understood why probabilities must be calculated in a particular way. Learners in the inquiry learning environment ended up in between: They equally understood the effect of variables differing within and between problem categories, but they scored lower on understanding why probabilities must be calculated in a particular way than learners in the self-explanation-based learning environment. For the most part, this is consistent with our hypotheses. Learners in the inquiry learning environment were expected to score the highest on conceptual knowledge, as developing and testing hypotheses by performing experiments clearly show underlying concepts and effects of variables on the probability of an event. Learners in this learning environment did so, along with learners in the self-explanation-based learning environment, which indicates that both performing experiments based on hypotheses and giving self-explanations afford the learners the opportunity to think about the underlying concepts instead of focusing only on the procedures, as is often the case in the domain of mathematics.

For intuitive knowledge, the results show that learners in the self-explanation-based learning environment acquired more intuitive knowledge than learners in the hypermedia learning environment and the observational learning environment, and learners in the inquiry learning environment ended up in between. Although this contrasts with expectations derived from the literature, as especially inquiry learning is supposed to lead to intuitive knowledge (Swaak & de Jong, 2001a), it matches the general pattern of results.

Differential effects for the four instructional approaches also appear on the procedural knowledge tests. Learners in the self-explanation-based learning environment performed the best: They knew how to calculate probabilities, and they were able to use this knowledge in both near and far transfer problems. Learners in the hypermedia learning environment also knew how probabilities should be calculated, but they were less able to show this in near or far transfer

problems. Learners in the observational learning environment scored low on procedural knowledge in general: They found it difficult to tell how probabilities should be calculated, and they had difficulties in solving near and far transfer problems compared to the other approaches. Finally, the inquiry learners found it difficult to tell how probabilities should be calculated, but they performed well on near transfer items and particularly well on far transfer items. These findings contrast with our hypotheses. First, we expected learners in the hypermedia learning environment to score highest on near and far transfer problems, as viewing the information in different representational codes would lead to a flexible and firm knowledge base. The fact that learners in the hypermedia learning environment did not score as high as expected (they even scored the lowest) can indicate that the learning processes supposed to be elicited by this instructional approach do not foster the ability to solve near and far transfer problems. Another explanation is that learners did not or were not able to identify their knowledge needs, which was one of the assumptions of this approach. Research in hypermedia learning has shown that learners with high prior knowledge search for information verifying what they already know, whereas learners with low prior knowledge hardly engage in elaborating strategies, such as making inferences by relating and integrating (Moos & Azevedo, 2008). If, for whatever reason, the learners do not identify their lack of knowledge, they will not search for more or new information, leaving them with a fragmented and incomplete knowledge base, which fails to lead to effective problem solving (cf. Scheiter et al., 2006). The short time spent by learners in this environment can be an indication that learners indeed only viewed a few information sources. Second, learners in the inquiry learning environment scored beyond expectations on procedural knowledge in general and on far transfer problems in particular. The finding was unanticipated, as inquiry learning focuses mainly on conceptual knowledge and less on procedural knowledge, but can be accounted for. In the domain of mathematics, conceptual and procedural knowledge are closely connected to each other (Rittle-Johnson, Siegler, & Alibali, 2001). An increase in conceptual knowledge directly leads to a gain in procedural knowledge (Rittle-Johnson & Alibali, 1999). Investigating the domain by conducting experiments and drawing conclusions ensures that learners develop a knowledge base in which the underlying concepts and relations between variables are firmly embedded. This conceptual insight into the subject matter is subsequently responsible for the ability to solve novel problems.

Finally, the picture for situational knowledge is the same as for the other types of knowledge: Learners in the self-explanation-based learning environment ended up with more situational knowledge than learners in the hypermedia learning environment and the observational learning environment, and the inquiry learners scored in between. This means that there were differences in learners' ability to identify relevant information in a given problem and to place the problem in the correct problem category. This finding is not consistent with our hypotheses. We expected learners in the observational learning environment to score highest on situational knowledge, as learners were

explicitly told about problem categories and why and when to use a specific strategy in a particular problem-solving situation. This approach assumes that learners build their cognitive representation of (the problem categories in) the domain by understanding the solution steps and the accompanying rationale, by relating and integrating this new knowledge to their prior knowledge, and by actively reflecting on the information given by the expert. The fact that learners in the observational learning environment did not score as high as expected can indicate that learners did not actively encode the information they received from the expert and were just passively watching the expert.

In the present study, the focus was on assessment of immediate effects of instruction. It would be interesting, though, to investigate to what extent the knowledge of the learners in the four different instructional approaches would be maintained over a longer period of time, for instance, after 6 months. In line with the results of studies such as Hulshof, Eysink, Loyens, and de Jong (2005) and Dean and Kuhn (2007), the higher achievement found in the present study for self-explanation-based learning and inquiry learning might be expected to be retained or even to become more apparent over a longer time period.

Assessing Instructional Approaches on Efficiency

We have seen that self-explanation-based learning was the most effective of the four instructional approaches. However, analysis of the times learners spent in the learning environments shows that self-explanation-based learning takes a lot of time, while hypermedia learning takes the least time. As a consequence, if we look at efficiency of learning (either Efficiency I or II), it can be concluded that hypermedia learning is the most efficient, while self-explanation-based learning is highly inefficient. Although learners in the hypermedia learning environment can decide which type of information they want to observe by selecting corresponding representations, the instructional material is presented to the learner in a more or less direct way. In the self-explanation-based learning environment, learners are prompted to self-explain the examples presented, which takes time. Given this line of reasoning, it is remarkable that no significant differences in efficiency were found between the observational learning environment in which learners watch an animation and the inquiry learning environment in which learners are supposed to generate hypotheses, perform experiments, and give explanations.

Generating Subject Matter

If we have a look at the results of our comparison, we can see a partition between on the one hand self-explanation-based learning and inquiry learning being more effective instructional approaches and on the other hand observational learning and hypermedia learning being less effective. If we relate this to the influential and disputed Kirschner et al. (2006) paper in which a distinction is made between approaches advocating problem-solving search and approaches advocating direct, explicit instruction, an interesting and

divergent picture appears. Where Kirschner et al. claimed that direct, explicit instruction is more effective than approaches that ask for problem-solving search, the present study questions this claim. The two most effective approaches can be characterized by the fact that the learners had to generate (parts of) the subject matter (i.e., searching a problem space) and the two less effective approaches had in common that the subject matter was more or less directly presented to the learners (i.e., direct instruction). According to Kirschner et al., searching a problem space in a highly complex environment makes heavy demands on working memory and is highly inefficient without contributing to learning. Indeed, the results of the present study show that learners that had to generate (parts of) the subject matter themselves experienced more cognitive load (intrinsic, extraneous, germane, and overall) and used more time than learners in the approaches in which the learning material was more or less directly presented. Learners apparently solve the temporary cognitive overload by spreading out their activities over time, making it less efficient. However, this does not mean, as claimed by Kirschner et al., that this has a detrimental effect on learning. On the contrary, the results show that the approaches in which the learners had to search a problem space in order to generate (parts of) the subject matter led to higher performance.

Summary of Results

In summary, the comparison of the four popular instructional approaches leads to four essential contributions to the field of instructional design of multimedia learning environments: (a) Having learners generate (parts of) the subject matter leads to better performance than presenting the subject matter, (b) this positive effect of generating (parts of) the subject matter influences all types of knowledge, (c) this positive effect is largest when having learners generate self-explanations in combination with worked-out examples; however (d) this positive effect is at the expense of efficiency of learning.

In the remainder of this discussion, the practical implications and the scope of results will be discussed. Furthermore, ways to increase the efficiency of self-explanation-based learning without losing effectiveness will be explored, as well as increasing the effectiveness of hypermedia learning and inquiry learning by including self-explanation prompts.

Practical Implications

Mayer (2004) argued that cognitive activity is what really promotes meaningful learning. Instructional designers should be aware of the fact that such cognitive activity does not necessarily occur spontaneously in practical situations, but that it needs to be stimulated. Approaches that take this into account by prompting learners to self-explain or stimulating them to hypothesize, experiment, and investigate in a simulation have now proved to be more effective than approaches that do not explicitly ask for cognitive

activity. To be more specific, our results indicate that if enough time is available in an educational setting, the best approach for instruction is self-explanation-based learning. This approach leads, overall, to the best learning results. Learners generate self-explanations and, as a result, acquire high-quality conceptual knowledge (including intuitive knowledge), as well as good procedural and situational knowledge. However, when time is scarce, hypermedia learning or inquiry learning are preferred. Hypermedia learning takes the least time, but instruction is more or less direct and learners will consequently focus mainly on conceptual knowledge, whereas intuitive knowledge, procedural knowledge (near as well as far transfer), and situational knowledge will lag behind. If a bit more time is available, the inquiry learning approach is a better option. Learners using this approach have to generate their own data by conducting experiments and they have to search the problem space to explain the pattern of results. As a result, they will end up with good conceptual, intuitive, and situational knowledge as well as good procedural knowledge, especially on far transfer.

Scope of Results

The main purpose of this article was to compare four modern instructional approaches. An instructional approach can be modeled in many different ways. Representations and cognitive tools must be chosen that will optimize the effects of the instructional approach. For the present study, a variety of learning environments was developed for each instructional approach. In the data analyses, all versions of the learning environments within each approach were taken together. This was possible, as analyses showed that there were no significant differences concerning the subtypes of knowledge as used in this study between the instructional variations within the instructional approaches. By doing this we assured that the results that were found can be attributed to the instructional approach in general instead of to one specific setup of that instructional approach. The results, however, were obtained in the domain of probability theory. Probability theory falls within the domain of mathematics, which differs from other domains such as the empirical sciences of biology, chemistry, or physics by its formal and abstract nature. Learners find it hard to understand that arithmetical representations stand for more than merely indicating which operations should be performed in order to come to the right solution. They are often not aware of the fact that the mathematical symbols represent principles and concepts underlying the procedures, and as a consequence they only focus on the procedural side of the domain (Cheng, 1999). This leads to an incomplete knowledge base consisting of procedural knowledge that is prone to errors, easily forgotten, and difficult to transfer to other problems (Ohlsson & Rees, 1991). Although empirical sciences have to contend with similar problems, further research should investigate whether the results found in the present study are also valid to other domains.

Making Self-Explanation-Based Learning More Efficient

We found that compared to the other instructional approaches, self-explanation-based learning is highly effective but inefficient. The question arises whether this instructional approach can be modified in such a way that it becomes more efficient without losing effectiveness. There are presumably two main factors contributing to long learning times: (a) the mere act of typing the written self-explanations (cf. Schworm & Renkl, 2006) and (b) generating high-level self-explanations (see Chi et al., 1989; Schworm & Renkl, 2007).

The typing requirement can be avoided by implementing the possibility of menu-based self-explanations in which the learner selects, for example, a certain self-explanation statement that justifies a solution step from a menu. Atkinson et al. (2003) found in two experiments that menu-based self-explanations actually fostered learning in probability theory without requiring additional learning time as compared to a condition without self-explanation requirement. Alevén and Koedinger (2002, Experiment 1) found that learners who were given the possibility of menu-based self-explanations in an intelligent tutoring system on geometry needed on average 18% more learning time as compared to a group without self-explanation requirement, a difference that was not significant.

Another possibility is to train mental self-explanations in advance. Beyond the time costs in the beginning, self-explanation training can induce effective but not time-consuming processing in subsequent learning environments. For example, Busch, Renkl, and Schworm (2008) found that learners who received a self-explanation training in the content area of fables provided better self-explanations and achieved better learning outcomes in scientific argumentation compared to an untrained control condition without taking additional learning time. In the long run, the initial time investment for training may pay off because remedial interventions get superfluous.

A final possibility of saving learning time in self-explanation-based learning is to simply restrict learning time. The self-explanation effect is robust even if learning time is held constant with reference to conditions without self-explanation requirements (e.g., Alevén & Koedinger, 2002, Experiment 2; Renkl, 1997). A question that, however, cannot be answered with certainty in this context is whether the present self-explanation-based approach would still be more effective than the other instructional approaches analyzed in this study if the learning time would have been restricted to the average level of those other instructional approaches. This issue should be solved by further experiments.

Combining Inquiry Learning With Self-Explanation Prompts

Instead of making self-explanation-based learning more efficient, we could also try to make the other instructional approaches more effective by including prompts to self-explain. Many studies, including a review study by Roy and Chi (2005), have shown that self-explaining is associated with deep

learning gains. Therefore, prompts to self-explain may also be beneficial within other instructional approaches. In the domain of probability theory, Gerjets, Scheiter, and Catrambone (2006) added self-explanation prompts to their hypermedia learning environment. In this case, however, prompting for self-explanations did not improve learning and even impaired learning when the individual events method (also used in the present study) was used. Adding self-explanation prompts to inquiry learning shows more promising results, though. Rittle-Johnson (2006) compared direct instruction with self-explanation prompts to inquiry learning with self-explanation prompts in the domain of mathematical equivalence problems. She found that prompting learners to self-explain under conditions of direct instruction and inquiry learning promoted conceptual knowledge and near and far transfer equally well. Although further research should provide evidence for this, this finding suggests that inquiry learning combined with self-explanation prompts would be the best mixture of instructional elements. The results of the present study imply that such a combination will lead to a complete knowledge base consisting of good conceptual and procedural knowledge and provide insight for solving novel problems.

Note

The research reported in this article took place in the context of the Dutch-German LEMMA cooperation (Learning Environments MultiMedia and Affordances). We would like to thank NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; Netherlands Organization of Scientific Research) and the DFG (Deutsche Forschungsgemeinschaft; German Research Community), who funded the project. We would also like to thank all other project members: Alexander Renkl and Rolf Schwonke from the University of Freiburg, Peter Gerjets and Katharina Scheiter from the Knowledge Media Research Center (IWM-KMRC) in Tübingen, and Jeroen van Merriënboer and Fred Paas from the Open University of the Netherlands.

References

- Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147–179.
- Atkinson, R. K., & Derry, S. J. (2000). Computer-based examples designed to encourage optimal example processing: A study examining the impact of sequentially presented, subgoal-oriented worked examples. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 132–133). Mahwah, NJ: Lawrence Erlbaum.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*, 774–783.
- Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*, 199–209.
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523–535.

- Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344–370.
- Bandura, A. (1976). *Social learning theory*. Upper Saddle River, NJ: Prentice Hall.
- Barab, S. A., Bowdish, B. E., Young, M. F., & Owen, S. V. (1996). Understanding kiosks navigation: Using log files to capture hypermedia searches. *Instructional Science*, *24*, 377–395.
- Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*, 345–363.
- Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, *101*, 70–87.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, *41*, 1069–1077.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.
- Burke, P. A., Etnier, J. L., & Sullivan, H. J. (1998). Navigational aids and learner control in hypermedia instructional programs. *Journal of Educational Computing Research*, *18*, 183–196.
- Busch, C., Renkl, A., & Schworm, S. (2008). Towards a generic self-explanation training intervention for example-based learning. In G. Kanselaar, V. Jonker, P. A. Kirschner, & F. J. Prins (Eds.), *Proceedings of the 8th International Conference of the Learning Sciences 2008*. Utrecht, the Netherlands: ICLS.
- Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *29*, 3–20.
- Cates, G. L., Skinner, C. H., Watson, T. S., Meadows, T. J., Weaver, A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based decision making: An evaluation of interspersing procedures. *School Psychology Review*, *32*, 601–616.
- Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory and Cognition*, *22*, 606–615.
- Cheng, P. C. H. (1999). Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education*, *33*, 109–130.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations—How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.
- Cognition and Technology Group at Vanderbilt. (1992). The Jasper series as an example of anchored instruction: Theory, program, description, and assessment data. *Educational Psychologist*, *27*, 291–315.
- Collins, A. (1991). Cognitive apprenticeship and instructional technology. In L. Idol & B. F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 121–138). Hillsdale, NJ: Lawrence Erlbaum.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum.
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347–362.

- Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*, 384–397.
- de Jong, T. (2005). The guided discovery principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 215–229). Cambridge, UK: Cambridge University Press.
- de Jong, T. (2006a). Computer simulations—Technological advances in inquiry learning. *Science, 312*, 532–533.
- de Jong, T. (2006b). Scaffolds for computer simulation based scientific discovery learning. In J. Elen & R. E. Clark (Eds.), *Dealing with complexity in learning environments* (pp. 107–128). London: Elsevier Science Publishers.
- de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*, 105–113.
- de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–202.
- Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Holt and Co.
- Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers. *Cognition and Instruction, 7*, 41–54.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*, 104–121.
- Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*, 360–370.
- Gijlers, H., & de Jong, T. (in press). Sharing and confronting propositions in collaborative inquiry learning. *Cognition and Instruction*.
- Hill, J. R., & Hannafin, M. J. (1997). Cognitive strategies and learning from the World Wide Web. *Educational Technology Research and Development, 45*, 37–64.
- Hulshof, C. D., Eysink, T. H. S., Loyens, S., & de Jong, T. (2005). ZAPs: Using interactive programs for learning psychology. *Interactive Learning Environments, 13*, 39–53.
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*, 83–93.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.
- Klahr, D., & Dunbar, K. (1988). Dual-space search during scientific reasoning. *Cognitive Science, 12*, 1–48.
- Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.
- Kolloffel, B., Eysink, T. H. S., de Jong, T., & Wilhelm, P. (in press). The effects of representational format on learning combinatorics from an interactive computer-simulation. *Instructional Science*.
- Lawless, K. A., & Brown, S. W. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science, 25*, 117–131.
- Linn, M. C., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science, 313*, 1049–1050.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, *41*, 226–232.
- Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, *13*, 125–139.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*, 14–19.
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*, 43–59.
- Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, *33*, 270–298.
- Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer-simulation for control-theory: Learning-processes and instructional support. *Journal of Research in Science Teaching*, *30*, 821–844.
- Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction*, *8*, 103–179.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, *84*, 429–434.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1–4.
- Paas, F., & van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, *35*, 737–743.
- Pietrzak, R. H., Cohen, H., & Snyder, P. J. (2007). Spatial learning efficiency and error monitoring in normal aging: An investigation using a novel hidden maze learning test. *Archives of Clinical Neuropsychology*, *22*, 235–245.
- Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*, 144–161.
- Quinn, J., & Alessi, S. (1994). The effects of simulation complexity and hypothesis generation strategy on learning. *Journal of Research on Computing in Education*, *27*, 75–91.
- Reiser, R. A., & Dempsey, J. V. (2002). *Trends and issues in instructional design and technology*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1–29.
- Rieber, L. P. (1993). Effects of visual grouping strategies of computer-animated presentations on selective attention in science. *Educational Technology, Research and Development*, *39*, 5–15.
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*, 1–15.
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*, 175–189.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*, 346–362.
- Rouet, J. F., Levonen, J. J., Dillon, A., & Spiro, R. J. (1996). *Hypertext and cognition*. Mahwah, NJ: Lawrence Erlbaum.
- Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 271–287). Cambridge, UK: Cambridge University Press.

- Scheiter, K., Gerjets, P., & Catrambone, R. (2006). Making the abstract concrete: Visualizing mathematical solution procedures. *Computers in Human Behavior, 22*, 9–25.
- Schmidgall, M., & Joseph, L. M. (2007). Comparison of phonic analysis and whole word-reading on first graders' cumulative words read and cumulative reading rate: An extension in examining instructional effectiveness and efficiency. *Psychology in the Schools, 44*, 319–332.
- Schnotz, W., & Hei, A. (2009). Semantic scaffolds in hypermedia learning environments. *Computers in Human Behavior, 25*, 371–380.
- Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*, 313–322.
- Schunk, D. H., & Hanson, A. R. (1989). Self-modeling and children's cognitive skills. *Journal of Educational Psychology, 81*, 155–163.
- Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology, 18*, 37–55.
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293–305.
- Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education, 46*, 426–445.
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*, 285–296.
- Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In T. Duffy & D. Jonassen (Eds.), *Constructivism and the technology of instruction* (pp. 85–107). Hillsdale, NJ: Lawrence Erlbaum.
- Spiro, R. J., & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia* (pp. 163–205). Hillsdale, NJ: Lawrence Erlbaum.
- Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: The development of the what-if test. *Studies in Educational Evaluation, 22*, 341–362.
- Swaak, J., & de Jong, T. (2001a). Discovery simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning, 17*, 284–294.
- Swaak, J., & de Jong, T. (2001b). Learner vs. system control in using online support for simulation-based discovery learning. *Learning Environments Research, 4*, 217–241.
- Swaak, J., van Joolingen, W. R., & de Jong, T. (1998). Supporting simulation-based learning: The effects of model progression and assignments on definitional and intuitive knowledge. *Learning and Instruction, 8*, 235–252.
- Tarmizi, R., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*, 424–436.
- van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26.
- van Gog, T., Paas, F., & van Merrinboer, 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 Merrinboer, 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 Merrinboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*, 5–14.

Learner Performance in Multimedia Learning

- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147–177.
- Wetzel, D. C., Radtke, P. H., & Stern, H. W. (1994). *Instructional effectiveness of video media*. Hillsdale, NJ: Lawrence Erlbaum.
- Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2006, April). *Observational learning from animated models: The relation between pacing, structure, and modality*. Paper presented at the Annual meeting of the American Educational Research Association, San Francisco.
- Wouters, P., Paas, F., & van Merriënboer, J. (2008). How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research, 78*, 645–675.
- Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology, 89*, 231–247.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3–21.
- Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–166.

Manuscript received March 5, 2008

Final revision received April 22, 2009

Accepted May 9, 2009