ELSEVIER

Learning and Instruction 16 (2006) 266–278

Learning and Instruction

www.elsevier.com/locate/learninstruc

Process support in learning tasks for acquiring complex cognitive skills in the domain of law

Rob J. Nadolski*, Paul A. Kirschner, Jeroen J.G. van Merriënboer

Educational Technology Expertise Centre, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

Abstract

Whole tasks for acquiring complex skills are often too difficult for novices. To solve this problem, *process support* divides the problem solving into phases, offers driving questions, and provides feedback. A multimedia program was used to teach sophomore law students ($N = 82$) to prepare and carry out a plea. In a randomised 2×2 design with the factors number of phases and availability of driving questions, students solving a task with fewer phases performed better and more efficiently than students exposed to more phases. Also, students receiving driving questions performed better, although not more efficiently than students not receiving such questions. The results indicate that whole tasks should be accompanied by process support, although task characteristics might restrict the benefits of driving questions in this support.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Support; Problem solving; Practicals

1. Introduction

A Multimedia Practical is a self-contained electronic learning environment which provides context-relevant practice to students for acquiring complex skills such as diagnosing a disease, selecting a job applicant, modelling stress factors that cause mental overload in workers, or preparing a plea to be held in court (Brown, Collins, & Duguid, 1989; Nadolski, Kirschner, van Merriënboer, & Hummel, 2001; Westera & Sloep, 1998). Such practicals provide an authentic setting for learners to develop the cognitive schemata necessary for acquiring complex skills whose mastery involves coordination and integration of its constituent skills and not simply the mastery of those separate constituent skills. Traditional Instructional Design models (e.g., Dick & Carey, 1979; Romiszowski, 1981) are not suitable for designing learning environments for acquiring such skills because they focus too strongly on the acquisition and training of constituent skills instead of the *complete* cognitive skill. Instructional Design models using whole-task approaches with Multimedia Practicals could overcome this shortcoming. Many researchers (e.g., Hannafin, Land, & Oliver, 1999; Jonassen, 1999; Mayer, 1999; Merrill, 2002; Stark, Gruber, Renkl, & Mandl, 1998; van Merriënboer, 1997) agree that transfer-oriented learning can best be achieved by using realistic *learning tasks*

* Corresponding author.
E-mail address: rob.nadolski@ou.nl (R.J. Nadolski).

0959-4752/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.learninstruc.2006.03.004

consisting of an authentic task description, an environment to carry out the task, and feedback on the quality of task execution. Such transfer-oriented learning is aimed at the mindful abstraction of knowledge from concrete learning experiences to be effectively used in a variety of previously unencountered settings.

This seemingly clear recommendation, however, is not easy to follow. Carrying out realistic whole tasks is often too difficult for the novice learner. A common solution is first to conceptually model reality (i.e., simplify it) and then pedagogically model that model (Achtenhagen, 2001). One way to pedagogically model the model, via didactic specification (Resnick, 1976), is to (a) *segment* the whole learning task into smaller task assignments, thus dividing the problem-solving process into *phases*, and (b) provide *driving questions* to guide carrying out the phases (Land, 2000). Feedback is then given at the completion of phases since its provision is inextricably intertwined with learning (Mory, 1996, 2003). *Process support* divides the problem solving into phases, offers driving questions, and provides feedback.

Problem solving, as operationalised in this research, involves higher-order thinking skills beyond those required for answering questions at the knowledge and comprehension level. It is a process where students discover a combination of previously learned rules and apply them in novel problem situations (Gagné, 1977; Lock, 1990). This process is augmented by domain-specific *process support* (see, e.g., Chinnappan & Lawson, 1996), which is tuned to the task at hand. Thus, process support is *not* directed at using content-free heuristics such as general problem-solving methods or content-free questions, for instance, asking if the person checked the spelling, or added an index when writing a report. Different problem characteristics require different problem-solving processes (Sternberg & Frensch, 1991) and therefore require different educational settings to learn to solve such problems (Jonassen, 1997, 2000). In such cases, the impact of process support on the learning results can be expected to be domain dependent. The present study is restricted to the domain of Law, which is characterised by convergent solutions and interdependent phases. Problem solving in domains with ontologies similar to that of Law such as philosophy and mathematics which are also argumentative and dialogical, probably require similar process support while problem solving in domains with other ontologies (e.g., inquiry-based science, or design-based engineering) and other problem characteristics (e.g., divergent solutions, independent phases) might require different process support.

One approach to segmentation is to present *process worksheets* (van Merriënboer, 1997), which provide distinct successive phases to the learners and guide them through the problem-solving process of the whole learning task. This process worksheet provides a systematic approach to problem solving for the whole learning task that should enhance problem-solving behaviour (e.g., Mettes, Pilot, & Roossink, 1981). Through tailoring the number of phases and providing feedback at the end of each phase, complex learning tasks come within reach of learners' capabilities. Whole tasks or tasks with too few phases are often too difficult and too mentally demanding for learners, preventing them from accurately processing the necessary information because they experience cognitive overload or revert to superficial, non-meaningful learning to keep their cognitive load within the threshold limit (Craik & Lockhart, 1972; Sternberg & Frensch, 1991). Tasks with too many phases may hamper learning because the redundant information between phases makes the information non-coherent, and/or the excess of details makes the tasks too mentally demanding (Mayer & Moreno, 2002). Finally, learners may regard the many phases as being too specific for the learning task in question, preventing them from constructing the generalisations or more abstract cognitive schemata necessary for learning transfer. In a previous experiment, we found that the number of phases in a whole learning task on Law affected task performance such that students receiving an intermediate number of phases outperformed students receiving either a low or a high number of phases (Nadolski, Kirschner, & van Merriënboer, 2005).

Driving questions are open questions given at the start of a phase that guide learners in how to carry out a phase. They do this, for example, by suggesting relevant procedures and principles, activating relevant prior knowledge, and/or referring to information resources in a correct and efficient way. Driving questions scaffold the problem-solving process within the phases of whole learning tasks. The benefits of such questions, combined with feedback, on reflection, exploration and self-directed learning are well documented (Dochy, 1992; Douglas, Hosokawa, & Lawler, 1988; Morgan & Saxton, 1991; Naidu & Bernard, 1992; Orlich, Harder, Callahan, Kauchak, & Gibson, 1994; Rowntree, 1992). In the present study, the driving questions primarily guided learners in how and when to use appropriate resources and how to select relevant elements from these resources. In other words, they were process oriented. Driving questions were provided at the start of the phase and kept available during task execution and could be answered by the student in an arbitrary order. They are expected to result in higher learner performance because they focus on higher-order skills essential in problem solving (see, e.g., Smith & Ragan, 1999). However, since tasks in the domain of Law have a convergent solution (i.e., as one proceeds in solving the problem the solution space decreases), the beneficial use of driving questions might be restricted to the early phases in problem solving.

Combined, the process worksheet focuses on the problem-solving process of the whole learning task, the driving questions on the problem-solving process within the phases, and the feedback at the end of each phase supports the judgements of the students during problem solving. In whole-task performance, an interaction effect might occur between the number of phases and the availability of driving questions. For a low number of phases, the problem-solving process associated with the large phases can be expected to be (too) difficult when driving questions are not provided. Driving questions may have a positive effect on completing the phases in this situation, and thus on the performance for the whole task. For a high number of phases, the problem-solving process for each phase will be simpler, so that the completion of the phases can be expected to be manageable without driving questions. As a consequence, if the whole task is divided in a low number of phases, providing driving questions is expected to be beneficial to the problem-solving process for each phase and to whole-task performance. If the whole task is divided in a high number of phases, giving driving questions might provide too much process support for each phase, to the eventual detriment of whole-task performance.

Each of the process support mechanisms used can, either separately or in combination, not only affect the effectiveness of the Multimedia Practical as reflected in the quality of the task performance, but also affect the efficiency of the task performance, defined here as task performance in relation to a combination of the mental effort, time on task, and motivation necessary to reach this level of task performance. In the remainder of this article task performance efficiency is abbreviated to task efficiency. Although efficiency can be operationalised in many different ways (see Admiraal, Wubbels, & Pilot, 1999; Kalyuga, Chandler, & Sweller, 1998; van Merriënboer, Schuurman, de Croock, & Paas, 2002), higher efficiency always indicates equivalent results with lower investment, higher results with the same investment, or, ideally, higher results with lower investment. In our study, investment is measured in terms of task motivation (Bonner, 1994; Maynard & Hakel, 1997), mental effort (Paas & van Merriënboer, 1994), and time on task (Karweit, 1984). Students are motivated if they find the task interesting to them and are willing to put a lot of energy into coming up with the best possible solution to the task (Maynard & Hakel). Perceived mental effort is a good indicator of how mentally demanding a task is (for an overview of studies, see Paas, Tuovinen, Tabbers, & van Gerven, 2003). If, for example, two students reach the same performance level with the same time on task and the same motivation, but with differing levels of mental effort, then the student experiencing the least mental effort is said to be more efficient. In a previous study (Nadolski, Kirschner, & van Merriënboer, 2005), this effect was found with respect to the number of phases. Students receiving a high number of phases were least efficient. We expect driving questions to have either positive or neutral effects on motivation and mental effort. They decrease the problem-solving space and demonstrate the relevance of the expected outcome to the learners. Although this might stimulate learners to become somewhat more motivated and invest more effort in learning, it is yet expected to decrease *overall* perceived mental effort because the process for carrying out the task is greatly facilitated, that is, learners have to invest much less effort in searching the problem space. Furthermore, we expect that the necessary time for answering driving questions also counterbalances the extra time needed for finding the information when no questions were given. In other words, driving questions should not result in longer time on task and thus be advantageous to efficiency, at least in the early phases of problem solving.

In doing such research, a number of potentially confounding variables need to be addressed. The first and possibly most important is differences between the participants with respect to prior knowledge (see, e.g., Dochy, 1992). One way of dealing with this issue is to make use of a randomised assignment of participants to experimental conditions. This has been done in the present study. Another way of dealing with this problem is to determine participant comparability with respect to prior knowledge before the experiment. In the present study, we are reasonably sure that participants will have comparable prior knowledge with respect to the domain of Law as they are all confronted with virtually identical first year Law curricula, being for all Dutch universities virtually identical with respect to both courses taken and textbooks used. For the same reason, we expect that participants will not have relevant prior knowledge with respect to pleading a case since the standardized curricula of Dutch Law Programmes throughout the country do not contain pleading a case in either their freshman or sophomore year. A second confounding variable might be student satisfaction with the instructional materials. If students regard the quality of the instructional material to be insufficient, they are less inclined to study it, which will also be reflected in their motivation (Keller, 1983). In the case of the research presented here, the basic material is the standard competency-based multimedia practical used in the Open University's Law programme and has been shown to be appealing (see, Nadolski, Kirschner, & van Merriënboer, 2005). A third issue that could confound the results of the experiment is whether students behave in the learning environment in the expected manner. Students are free to control their own learning to a large extent in a Multimedia Practical so that the study and problem-solving behaviour is inextricably bound up with this shift of learning control (Chin, 1997). In the research presented here, all student behaviour is logged such that possible differences in behaviour can be noted (see, e.g., Land, 2000). This technique was also used in earlier studies (Nadolski et al., 2001; Nadolski, Kirschner, & van Merriënboer, 2005). In those studies, all students acted in an expected and similar manner. Finally, since the learning materials in this study are completely computer delivered, students' attitude towards learning with computers as well as computer literacy can confound the results. To control for this, students answered a number of questions designed to determine whether such differences existed.

The present study is conducted in an ecologically valid setting and employs a randomised 2×2 factorial design to examine the effects of the number of phases (low, high) and driving questions (present, absent) on the performing of law tasks and on the efficiency of that performance. The low number of phases in this study is comparable to the intermediate number of phases in the aforementioned study (Nadolski, Kirschner, & van Merriënboer, 2005) and the high number of phases is the same in both studies. The first hypothesis is that students solving a whole learning task with a low number of phases will exhibit higher performance and be more efficient learners than students exposed to a high number of phases. The second hypothesis is that students receiving driving questions for each phase will exhibit higher performance and will be more efficient than those not receiving driving questions. The third hypothesis pertains to a possible interaction between the number of phases and the availability of driving questions namely that driving questions have little added value if the number of phases is high, but become more valuable if the number of phases is low.

2. Method

2.1. Participants

Eighty-two sophomore law students from five Dutch universities (49 females, 33 males; mean age = 23.5 years, SD = 4.2) took part in the experiment and were at each university randomly assigned to one of four conditions. The conditions were low number of phases without driving questions ($n = 21$); low number of phases with driving questions ($n = 18$); high number of phases without driving questions ($n = 22$); high number of phases with driving questions ($n = 21$). None of the participants had prior plea experience.

2.2. Learning materials

The Multimedia Practical *Preparing a Plea* (Wöretshofer et al., 2000) was adapted for this experiment. In the practical, the learner is a trainee in a virtual law firm. The trainee first studies a general introduction to pleading a case in which supportive information and various support tools are provided. Supportive information is helpful to the learning and execution of problem-solving aspects of learning tasks. This is often regarded as 'the theory' by teachers. The support tools include video examples of lawyers conducting a plea, discussions of ethical issues in pleading a case, numerous tips on the communicative aspects in pleading a case, and judicial procedural aspects of plea preparation. During this general introduction, the trainee receives several assignments to guide the study of the theory as well as support from a senior (virtual) employee of this firm, the coach. The trainee can make use of standard office equipment and can visit other places in the firm. The trainee can, for example, study the legal backgrounds of different cases in a file cabinet, observe and analyse other pleas using a "plea checker", make electronic notes, attend staff meetings, and consult experts. After this general introduction, the trainee must prepare pleas for various cases (i.e., whole tasks). The case files are available within a (virtual) office. The coach provides task assignments with feedback for each segment in the whole task. Finally, the trainee conducts the prepared pleas outside the Multimedia Practical in an actual simulated courtroom. In adapting this practical for the experiments, teachers ($n = 6$) were used for the identification of functional and non-trivial phases, so that all phases identified were considered to be real life and non-artificial (Nadolski et al., 2001).

Four versions of the practical were produced for the different experimental conditions. The complexity of the assignments (1 = very simple, 2 = simple, 3 = complex, 4 = very complex) for all versions of the practical was determined by 32 experts (20 teachers, 12 graduate Law students). The complexity of the assignments was measured with the aid of the task-complexity instrument developed by Nadolski, Kirschner, van Merriënboer, and Wöretshofer (2005). The mean complexity for the tasks was 2.7 (SD = 0.3) for the low number of phases − no driving questions; 2.3 (SD = 0.3) for the low number of phases − driving questions; 2.0 (SD = 0.4) for the high number of phases − no driving questions; and 1.9 (SD = 0.4) for the high number of phases − driving questions. All versions of the practical contained the same general introduction to pleading a case with identical supportive information and support tools. The goal of the practical was to prepare students to carry out a plea in court, a task that was indicated with a mean complexity of 3.7 (SD = 0.7) by the aforementioned 32 participants. Fig. 1 provides an overview of the learning materials for the four conditions and should be viewed in connection with Table 1.

Participants received a whole-task training using three learning tasks, one of which was compulsory, plus a task to determine transfer. All versions contained case files and legal documentation (i.e., sections of law codes, jurisprudence), which were available to the participants while they were working on the learning tasks and the transfer task. The material for all learning tasks (thus not the transfer task) also contained process worksheets that differed for each condition. Final performance on the whole learning task (i.e., the plea) is considered proof of skill acquisition. For all learning tasks, each phase of the process worksheet was accompanied by a task assignment, with or without driving questions and the feedback to each assignment was provided at the end of each phase. In the case of driving questions, this feedback was also specific to those driving questions. For all learning tasks, all conditions included condition-specific task assignments with feedback provided at the end of each phase. As the task assignments do not have univocal answers, the feedback typically represents expert model(s). The feedback supports the self-judgement by students. One of the phases results in a *plea inventory*: these are the selected elements from the case file and legal documentation that might be usable in a plea (see Table 1). The phase in which a plea inventory had to be constructed was included in all conditions. In the conditions with a low number of phases it was the first phase; in the conditions with a high number of phases it was the second phase (see Fig. 1). The material for the transfer task was the same in all conditions and did not include phases, driving questions, or feedback. In other words, no process support was given for the transfer task. Final performance on the transfer task (i.e., the plea) is considered proof of skill transfer.

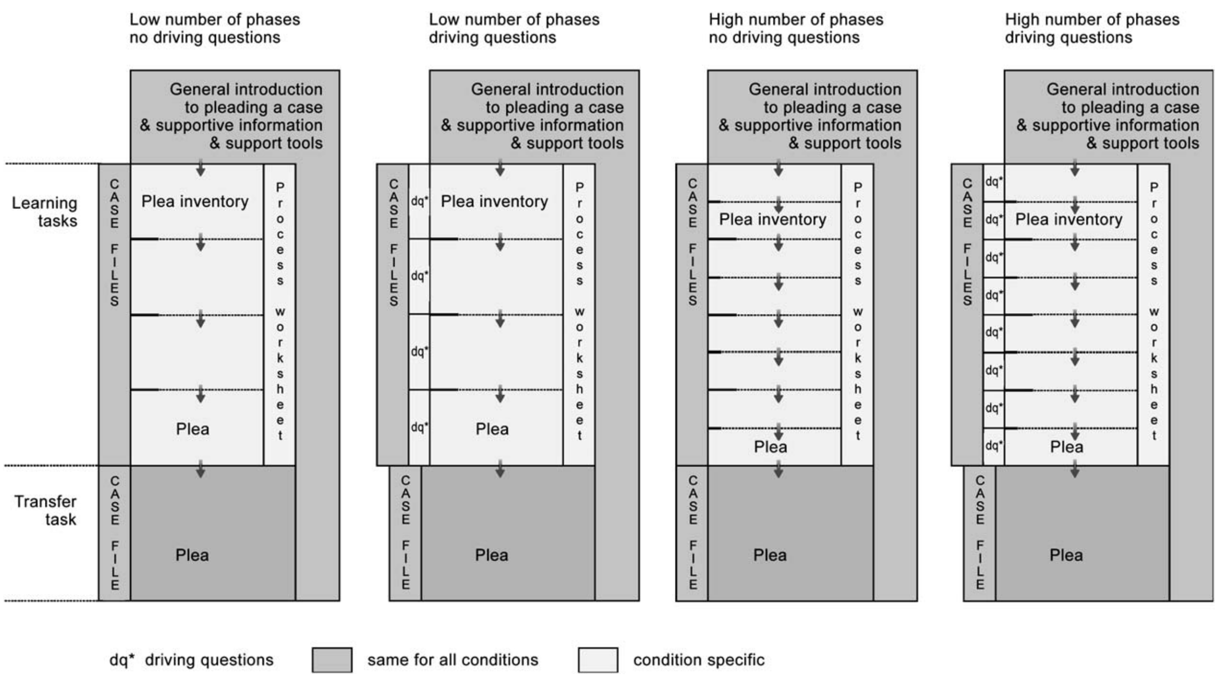

Fig. 1. Overview of the learning materials in the four conditions.

Table 1
Driving questions for 'get acquainted with a law case': the answers result in a plea inventory

What field of law does this case belong to?
Who are the parties in this case?
Whose representative will you be?
Who is bringing the case before the court?
What kind of procedure should be used?
What kind of judge will try this case? Does the judge have jurisdiction?
What about the territorial jurisdiction of the judge?
What is this case roughly about?
Is there a previous judicial/legal history, and if so, is this relevant for this case?
What do you already know about the subject matter in this case?
Is the file complete?
What legal aspects could be of importance in this case?
What clues does the case offer at first sight?
What is the tentative goal of your plea?

2.3. Measurement instruments

2.3.1. Background questionnaire

A background questionnaire was used to gather data on prior plea related experience (i.e., writing skills (six-item, 4-point scale) and oral presentation skills (six-item, 4-point scale), debating club membership (dichotomously)), computer literacy (six-item, 4-point scale), attitude towards learning with computers (six-item, 4-point scale), age, and gender.

2.3.2. Performance instrument for whole learning task

The quality of the plea in the compulsory learning task was based upon an evaluation of the videotaped participants' pleas (on a 7-point scale). This final rating was based on inclusion of relevant content (39 items), and coherence of the learning task plea (seven items). Examples of coherence are 'starts with a conclusion to be drawn from the plea', 'provides a content overview of the plea before going into details', 'indicates how successive arguments relate to relevant facts', 'various arguments are treated separately'. The performance for the learning task plea was scored on those 46 items and this was then transformed into a grade on the 7-point scale (Cohen's Kappa = 0.6). This same instrument also proved to be reliable and content valid in a previous experiment (Cohen's Kappa = 0.6; see Nadolski, Kirschner, & van Merriënboer, 2005).

2.3.3. Performance instrument for whole transfer task

The quality of the plea in the transfer task was based upon an evaluation of the videotaped participants' pleas (on a 7-point scale). The items were similar to those mentioned in the former instrument. The final rating was based on inclusion of relevant content (46 items) and the coherence of the transfer task plea (nine items). The performance for the transfer task was scored on those 55 items and this was then transformed into a grade on the 7-point scale (Cohen's Kappa = 0.6). This instrument also proved to be reliable and content valid in a previous experiment (Cohen's Kappa = 0.6; see Nadolski, Kirschner, & van Merriënboer, 2005).

2.3.4. Performance instrument for plea inventory

The quality of the plea inventory (i.e., the inclusion of relevant content) for the compulsory learning task was based upon an evaluation of the written participants' plea inventory (on a 4-point scale). The performance for the plea inventory was scored on 38 items and this was then transformed into a grade on the 4-point scale (Cohen's Kappa = 0.7).

2.3.5. Instrument for time on task

Participants reported their time on task for each phase on a prestructured time sheet. They were instructed to note the start times and end times for each of the phases and to note the time spent in multiples of 5 min.

2.3.6. Mental effort rating scale

Participants indicated their mental effort for each phase on a 9-point rating scale (1 = very, very low mental effort, 9 = very, very high mental effort), developed by Paas, van Merriënboer, and Adam (1994). This scale was used to measure the perceived cognitive load of each of the constituting phases of the learning task, and the cognitive load of the transfer task (see Paas et al., 2003, for an overview of studies in which the same instrument has been used). Perceived mental effort was used as an indication for cognitive load; the less mentally demanding the task, the lower the cognitive load.

2.3.7. Motivation rating scale

Participants indicated their motivation for each phase on a three-item, 7-point rating scale (1 = very low motivation, 7 = very high motivation) that was developed by Maynard and Hakel (1997) (alpha = 0.83). Items were 'I was motivated to perform well on this task assignment', 'This task assignment was interesting to me', 'I put a lot of effort into coming up with the best possible solution'.

2.3.8. Satisfaction/perceived efficacy rating scale

Participants indicated their satisfaction with the feedback on an eight-item, 4-point scale (1 = very dissatisfied, 4 = very satisfied) (alpha = 0.75). Adequacy of the number of phases was indicated on a single 7-point scale (1 = far too few phases, 4 = perfectly all right, 7 = far too many phases). Whether participants felt that they had achieved the goal of the practical ('I've learned to conduct a plea') was indicated on a single 7-point scale (1 = very strongly disagree, 7 = very strongly agree). Relevance of the supportive information was indicated on a single 4-point scale (1 = strongly irrelevant, 4 = strongly relevant).

2.3.9. Computer logging

Since all conditions were computer supported, study behaviour was logged and analysed. Participants' logging files were electronically collected for this purpose.

2.4. Procedure

Participation in the experiment was equivalent to enrolment in the course. Information about the rigour of the practical (60 study hours in one month) and the required prior knowledge and skills were available before enrolment. Before taking part, participants were also informed that data would be gathered for scientific research and during enrolment gave permission to use the data for this purpose. Participants were randomly assigned to one of the four experimental conditions and were required to work individually. All materials were sent to the participants' home addresses or could be collected at their faculty addresses. All versions of the learning materials first presented the same general introduction to pleading a case with identical supportive information and support tools. After that, participants could work on the learning tasks. They were strongly advised to work phase by phase since the program did offer the possibility of skipping consecutive phases. Logging results indicated that the participants did not skip any phases. There was maximum learner control within a phase so that participants were free to decide if and when to consult phase-specific information and how long to work on the task assignment in a phase.

After two weeks (approximately 30 study hours), participants were required to make their plea for the compulsory learning task. This plea was videotaped for later evaluation. After working on the learning task(s), participants had access to the transfer task, which did not contain any support. About two weeks later, they were required to make their transfer task plea, which was also videotaped. There were strict time constraints for the pleas. As it is legally required for lawyers in the Netherlands to submit a plea note when pleading a case in court, participants were also required to include a plea note (i.e., a memorandum of oral pleading) for both tasks. Participants were informed that both the work that they carried out in both the compulsory learning task and the transfer task (i.e., their pleas and results from other assignments such as the plea inventory and the plea note) were going to be used for summative evaluation. Finally, they were required to return the completed background questionnaire, time sheet, and rating scales in a stamped self-addressed envelope along with electronically sending their logging results one week after completion of the course. Participants were informed whether they did or did not earn the study/course credits.

2.5. Data analysis and scoring

The experimenters extracted the plea inventory from the logging results and analysed logged study behaviour. Two judges blindly and independently scored all participants' word processed plea inventories and their videotaped pleas using the performance measurement instruments. All efficiency measures were calculated using a procedure that closely resembles the procedure described by Tuovinen and Paas (2004) for determining instructional condition efficiency.

Efficiency was calculated as

$$\frac{(P-E-T-Mv)}{\sqrt{(4)}}$$

where P = performance, E = mental effort, T = time on task, and Mv = motivation.

The P, E, T, and Mv scores on all variables are standardized (the total mean was subtracted from each score and the result was divided by the standard deviation), giving z-scores for each variable. This score can be negative, in the case that the sum of normalized scores for mental effort, time on task and motivation is greater than the normalized score for the performance. Although a negative group score mean indicates a lower value for efficiency than a positive group score mean, this does not necessarily mean that a negative group score mean is less efficient than a positive group score mean, it depends on the distance between the two group score means and the values for the standard deviation. Furthermore, there can also be a significant difference between two positive group score means or between two negative group score means.

3. Results

The collected data for determining computer literacy and attitude towards learning with computers showed no differences between conditions. First, none of the participants had prior plea experience, which was concluded from the collected background questionnaire data for prior plea experience. Second, participant comparability with respect to domain knowledge was assured because first year Law curricula of all Dutch universities are virtually identical with respect to both courses taken and textbooks used, and all participants were only in their sophomore year. Finally, the randomised set-up of this study ruled out the possibility of a prior knowledge domain knowledge effect if any differences remained. Analysis of logging results showed that the conditions did not differ with respect to the use of either the supportive information or the support tools included in the general introduction to pleading a case, spent comparable study time on the general introduction (approximately 15 study hours), etc., indicating that the study behaviour for all groups was similar and as expected. Analysis of logging results also indicated that all participants did not skip any phases during whole-task execution and did not spend any time on the non-compulsory learning tasks.

3.1. Performance

The mean performance results for the learning task, the transfer task, and the plea inventory are summarised in Table 2.

Table 2
Performance on the learning task, transfer task, and plea inventory

	Low number of phases				High number of phases			
	No questions ($n = 21$)		Questions[a] ($n = 18$)		No questions ($n = 22$)		Questions ($n = 21$)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Learning task* (1–7)	6.19	0.87	5.94	0.73	5.41	0.96	5.33	1.24
Transfer task (1–7)	3.29	0.85	3.39	0.61	3.19	0.51	3.14	0.73
Plea inventory** (1–4)	1.90	0.62	2.38	0.50	1.86	0.56	2.57	0.68

*$p < 0.01$ for the number of phases; **$p < 0.001$ for availability of driving questions.
[a] Only 16 of 18 plea inventory results were collected for this condition.

With regard to the learning task, ANOVA revealed a significant main effect for the number of phases on the performance of the learning task plea, $F(1, 78) = 10.39$, MSE = 0.95, $p < 0.01$, $\eta^2 = 0.12$. Students in the conditions with less phases ($M = 6.08$, SD = 0.16) significantly outperformed students in the conditions with more phases ($M = 5.37$, SD = 0.15) (max = 7). There was no main effect for driving questions and there was no interaction.

With regard to the transfer task, results on the performance of the transfer task plea revealed neither significant main effects nor interaction effects.

With regard to the quality of the plea inventory, ANOVA revealed a main effect for driving questions, $F(1, 76) = 19.02$, MSE = 0.36, $p < 0.001$, $\eta^2 = 0.20$. The conditions with driving questions ($M = 2.47$, SD = 0.61) significantly outperformed the conditions without driving questions ($M = 1.88$, SD = 0.59) (max = 4). There was no main effect for the number of phases and there was no interaction.

3.2. Time on task, mental effort, and motivation

The mean results for time on task, mental effort, and motivation are summarised in Table 3.

With regard to the learning task, there was a marginally significant effect for the number of phases on time on task, $F(1, 78) = 3.81$, MSE = 152045, $p < 0.1$, $\eta^2 = 0.05$. The conditions with less phases spent less time on task ($M = 940$ min, SD = 306) than conditions with more phases ($M = 1107$ min, SD = 449). There were no effects for the number of phases on mental effort and motivation. There was also no main effect for driving questions, and no interaction effects on time on task, mental effort and motivation.

With regard to the transfer task and the plea inventory, there were no significant main effects and no interaction effects on either of the dependent variables.

In sum, ANOVAs revealed neither significant main effects nor interaction effects on mental effort or motivation. Participants reported an average mental effort, ranging between 5.00 and 6.20 (max = 9) for all three tasks in all conditions. Participants in all conditions were highly motivated when working on the learning task and the transfer task, and preparing their plea inventory, with motivation scores ranging between 5.29 and 5.89 (max = 7).

3.3. Task efficiency

The mean efficiency results for learning task, learning plea, and plea inventory are summarised in Table 4. As mentioned earlier, efficiency was calculated as $(P - E - T - Mv)/\sqrt{(4)}$, where P = performance, E = mental effort, T = time on task, and Mv = motivation. The P, E, T, and Mv scores on all variables were standardized (the total mean was subtracted from each score and the result was divided by the standard deviation), giving z-scores for each variable.

Table 3
Time on task (in minutes), mental effort and motivation on learning task, transfer task, and plea inventory

	Low number of phases				High number of phases			
	No questions ($n = 21$)		Questions ($n = 18$)		No questions ($n = 22$)		Questions ($n = 21$)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Learning task								
Time on task	933	274	947	348	1048	453	1170	446
Mental effort (1–9)	5.77	1.03	5.35	1.15	5.54	0.94	5.45	0.88
Motivation (1–7)	5.34	0.83	5.29	0.79	5.51	0.70	5.64	1.00
Transfer task								
Time on task	351	177	360	174	425	309	411	213
Mental effort	5.70	1.74	5.42	1.27	6.20	1.45	5.48	1.41
Motivation	5.45	0.78	5.57	0.88	5.89	0.69	5.64	0.84
Plea inventory								
Time on task	210	84	208	115	175	103	172	77
Mental effort	5.19	1.57	5.00	1.88	5.14	1.32	5.10	1.22
Motivation	5.57	0.82	5.46	0.98	5.61	0.86	5.41	0.89

Notes: Time on task based upon self-report. Mental effort is measured on a 9-point rating scale (Paas et al., 1994). Motivation is measured on a three-item 7-point rating scale (Maynard & Hakel, 1997).

Table 4
Efficiency of the learning task, transfer task, and plea inventory

	Low number of phases				High number of phases			
	No questions ($n = 21$)		Questions ($n = 18$)		No questions ($n = 22$)		Questions ($n = 21$)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Learning task*	2.81	7.41	2.59	8.94	−0.43	11.64	−3.58	11.17
Transfer task	1.23	6.40	1.22	6.13	−1.63	10.95	−0.79	7.15
Plea inventory	−1.24	4.57	−0.72	6.43	0.58	5.27	1.28	4.14

*$p < 0.05$ for the number of phases.

With regard to learning task efficiency, ANOVA revealed a main effect for the number of phases, $F(1, 78) = 4.51$, MSE = 99.92, $p < 0.05$, $\eta^2 = 0.06$. The conditions with a lower number of phases ($M = 2.71$, SD = 8.04) were more efficient than the conditions with a higher number of phases ($M = -1.97$, SD = 11.38). There was no main effect for driving questions and no interaction effect.

There were neither significant main effects nor interaction effects for transfer task efficiency.

On the plea inventory efficiency, contrary to our expectations, ANOVA did not reveal a significant main effect for driving questions. There was no main effect for the number of phases and no interaction.

3.4. Satisfaction/ perceived efficacy

There was no main effect for number of phases and no interaction on user satisfaction with feedback (see Table 5). ANOVA indicated a significant main effect for driving questions, $F(1, 78) = 4.24$, MSE = 0.18, $p < 0.05$, $\eta^2 = 0.05$. Those in conditions with driving questions ($M = 3.31$, SD = 0.41) reported being significantly more satisfied with feedback than those where there were no driving questions ($M = 3.12$, SD = 0.43).

Finally, there were no significant main or interaction effects for adequacy of the number of phases, opinions on efficacy (‘I’ve learned to conduct a plea’), and opinions on the relevance of the supportive information.

4. Discussion

This study examined the effect of the number of phases and the availability of driving questions on both task performance and task efficiency. The results reported here show a main effect for the number of phases on both learning task performance and efficiency. A lower number of phases in learning to solve complex whole tasks led to both higher performance and greater efficiency. There were no differences between the conditions for transfer task performance and efficiency. These findings are in line with earlier research in which only the number of phases was varied (Nadolski, Kirschner, & van Merriënboer, 2005), where the low number of phases in the current experiment is comparable to the intermediate number of phases in the previous one and the high number of phases is the same as in the previous one. That transfer task performance and efficiency were not influenced by the number of phases is probably due to the fact that participants did not make use of the non-compulsory learning tasks provided. These tasks that were meant to

Table 5
Satisfaction and perceived efficacy of the training

	Low number of phases				High number of phases			
	No questions ($n = 21$)		Questions ($n = 18$)		No questions ($n = 22$)		Questions ($n = 21$)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Feedback*	3.05	0.44	3.24	0.41	3.18	0.42	3.38	0.40
Phase adequacy	5.20	1.32	4.78	0.94	5.18	0.73	5.33	0.97
Goal of practical	5.52	0.93	5.39	1.14	5.72	0.94	5.57	0.87
Supportive info	2.75	0.55	2.61	0.70	2.95	0.49	2.76	0.70

*$p < 0.05$ for availability of driving questions.
Notes: The scales were as follows: satisfaction with feedback (1 = very dissatisfied, 4 = very satisfied); phase adequacy (1 = far too few phases, 4 = perfectly all right, 7 = far too many phases); goal of practical; the assertion ‘I’ve learned to conduct a plea’ (1 = very strongly disagree, 7 = very strongly agree); supportive info: the assertion ‘the relevance of the supportive information’ (1 = strongly irrelevant, 4 = strongly relevant).

offer variability of practice were not used for two possible reasons. First, participants might have used their experience with the study time on the compulsory learning task as a reference, and thus might have regarded the execution of these tasks to cost too much study time, that is in excess of the scheduled 60 study hours. Second, they might have acted as a good ‘homo economicus’ always does, meaning that they only do what they are required to do in order to maximize their own well being (Hirsch, Michaels, & Friedman, 1990). In this case this means that assessment drives learning and since formal, summative assessment was limited to the compulsory learning task (i.e., one single whole learning task) and the transfer task, they did the minimum amount of work for achieving this. Researchers agree that transfer cannot be expected in such circumstances as there is not enough practice for the necessary schema acquisition (Spiro, Coulson, Feltovich, & Anderson, 1988), too little variability in practice (Paas & van Merriënboer, 1994), and not enough stimulation of mindful abstraction (Perkins & Salomon, 1989).

The results also show a positive main effect for driving questions on the plea inventory performance, which is one of the phases in the whole learning task. The availability of driving questions did not – contrary to our expectations – beneficially affect the efficiency of the plea inventory due to an unexpected greater time on task. Although participants who had to proceed through more phases showed – as expected – a trend towards spending more time on the whole learning task, an opposite trend was found for time on task on the plea inventory. This could be due to the position of the plea inventory assignment in the whole set of instructional materials. In conditions with fewer phases, the plea inventory is the first assignment, while in conditions with more phases it is the second assignment. Participants receiving the plea inventory as the first assignment might be inclined to invest more time than those receiving it as the second assignment since, for them, it is their first opportunity to get acquainted with the whole task. The absence of an effect of driving questions on both learning task performance and efficiency can be due to task characteristics and feedback given at the completion of phases. In our study, the problem-solving process can be seen as a sequence of interdependent phases converging on a solution for the whole learning task. In fact, as feedback is provided after each phase, each subsequent phase decreases the solution space for the whole task. Therefore, positive effects of driving questions are expected to extinguish in subsequent phases. This explanation, in addition to the observation that participants did not use the non-compulsory learning tasks, probably also accounts for driving questions not positively affecting transfer task performance and efficiency.

No interaction effects were observed between the number of phases and providing driving questions. This may be due to the relatively small sample size. In conjunction with a modest, probably too small, difference between the low and the high number of phases this might have prevented the expected added value of driving questions to occur. A second reason could be the previously mentioned decreasing solution space. An ethical note needs to be made here. As was already mentioned in the introduction, this experiment was carried out in an ecologically valid ‘real’ learning situation. Optimisation of research results by inclusion of a poor learning condition (i.e., where one can be fairly sure that learning is suppressed, for example, by omitting feedback after completion of phases) was not an option here. The learning materials with which we began could be considered to be ‘good’ and all of our four experimental conditions had the function of ‘making the material better’.

Finally, the participants expressed satisfaction with the instructional materials and in general their satisfaction with the quality of the instructional materials did not differ between the conditions, with the exception of those receiving driving questions who were more satisfied with feedback than those not receiving driving questions. This difference might be the result of these participants seeing a clear match between each driving question and its feedback. The collected motivation data, however, show that this did not affect their inclination to study the material and this did not affect the findings on the number of phases and the provision of driving questions.

The results obtained here give rise to several directions for future research, all of which should take place with larger sample sizes. First, a set of varied learning tasks for practice could be made compulsory, making sure that participants work on a set of learning tasks with high variability. Various studies have shown the benefits of such high variability practice on transfer (Cormier & Hagman, 1987; Paas & van Merriënboer, 1994; Quilicy & Mayer, 1996; Shapiro & Schmidt, 1982; Singley & Anderson, 1989). Second, future experiments must further explore the conditions under which driving questions lead to better performance and efficiency. These conditions appear to be related to whole-tasks characteristics such as the size of the problem space (Newell & Simon, 1972), task complexity, kind of solution (convergent vs. divergent), relationship between phases (interdependent vs. independent), and provision or omission of feedback after completion of the phases. For example, a study examining the effects of driving questions within generative and creative brainstorming tasks could provide more insight into the value of driving questions since the solution space is much larger. Third, future studies must take into account that process support may be domain related (e.g., Amsel, Langer, & Loutzenhiser, 1991). It is expected that the findings here can be extended to a domain with an ontology similar to that of Law. However, if the ontology of a domain is different, it is unclear if the findings of the current study could be replicated.

A straightforward practical implication of this study is that process support should be provided for whole learning tasks. One should split the whole task in a limited number of phases but not too many phases and provide feedback at the completion of phases. Practical implications for the use of driving questions are less clear. In this study, it was found that driving questions were only beneficial for early phases in problem solving, probably due to task characteristics. Therefore, more articulated research with respect to driving questions is needed. A final consideration is that the instructional material used presupposes that students have roughly the same prior knowledge and skills. However, mental effort and time on task may provide good input for tailoring the instructional materials to individual students while they are working with it via dynamic task selection. For instance, Salden, Paas, Broers, and van Merriënboer (2004) claim that the best new learning task for a person who is reaching a high level of performance with a very *low* investment of mental effort will be different from the best new learning task for a person who is reaching the same level of performance with a very *high* investment of mental effort. Such approaches to personalized, student-centered instruction not only provide whole tasks to students, but also offer the opportunity to build Multimedia Practicals that adapt their level of process support to individual learners.

Acknowledgements

The authors would like to thank the law students from various Dutch universities for their participation in this study and their teachers for letting them participate.

References

Achtenhagen, F. (2001). Criteria for the development of complex teaching–learning environments. *Instructional Science*, *29*, 361–380.

Admiraal, W., Wubbels, T., & Pilot, A. (1999). College teaching in legal education: Teaching method, students' time-on-task, and achievement. *Research in Higher Education*, *40*, 687–704.

Amsel, E., Langer, R., & Loutzenhiser, L. (1991). Do lawyers reason differently from psychologists? A comparative design for studying expertise. In R. J. Sternberg, & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 223–250). Hillsdale, NJ: Lawrence Erlbaum.

Bonner, S. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society*, *19*(3), 213–234.

Brown, J. S., Collins, A., & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*, 32–42.

Chin, C. (1997). Promoting higher cognitive learning in science through a problem-solving approach. React, issue 1. Retrieved July 14, 2004, from <http://eduweb.nie.edu.sg/REACTOld/1997/1/2.html>.

Chinnappan, M., & Lawson, M. J. (1996). The effects of training in the use of executive strategies in geometry problem solving. *Learning and Instruction*, *6*, 1–17.

Cormier, S. M., & Hagman, J. D. (Eds.). (1987). *Transfer of learning: Contemporary research and applications*. San Diego, CA: Academic Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Dick, W., & Carey, L. (1979). *The systematic design of instruction*. Glenview, IL: Scott Foresman.

Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning*. Utrecht, Netherlands/London, UK: LEMMA/Jessica Kingsley.

Douglas, K., Hosokawa, M., & Lawler, F. (1988). *A practical guide to clinical teaching in medicine*. New York: Springer.

Gagné, R. M. (1977). *The conditions of learning* (4th ed.). New York: Holt, Rinehart, & Winston.

Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory*, *Vol. 2* (pp. 115–140). Mahwah, NJ: Lawrence Erlbaum.

Hirsch, P. M., Michaels, S., & Friedman, R. (1990). Clean models versus dirty hands: Why economics is different from sociology. In S. Zukin, & P. DiMaggio (Eds.), *Structures of capital: The social organization of economy* (pp. 39–56). Cambridge, UK: Cambridge University Press.

Jonassen, D. H. (1997). Instructional design model for well-structured and ill-structured learning outcomes. *Educational Technology Research and Development*, *45*, 65–95.

Jonassen, D. H. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory*, *Vol. 2* (pp. 215–239). Mahwah, NJ: Lawrence Erlbaum.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*, 63–85.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, *40*, 1–17.

Karweit, N. (1984). Time-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership: Journal of the Association for Supervision and Curriculum Development*, *41*, 32–35.

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 383–434). Hillsdale, NJ: Lawrence Erlbaum.

Land, S. M. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, *48*, 61–78.

Lock, R. (1990). Open-ended, problem-solving investigations. *School Science Review*, *71*, 63–72.

Mayer, R. E. (1999). Designing instruction for constructivist learning. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory*, *Vol. 2* (pp. 141–159). Mahwah, NJ: Lawrence Erlbaum.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, *12*, 107–119.

Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance*, *10*(4), 303–330.

Mettes, C. T. W., Pilot, A., & Roossink, H. J. (1981). Linking factual knowledge and procedural knowledge in solving science problems: A case study in a thermodynamics course. *Instructional Science*, *10*, 333–361.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology, Research and Development*, *50*, 43–59.

Morgan, N., & Saxton, J. (1991). *Teaching, questioning and learning*. London: Routledge.

Mory, E. H. (1996). Feedback research. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 919–956). New York: MacMillan Library Reference.

van Merriënboer, J. J. G. (1997). *Training complex cognitive skills*. Englewood Cliffs, NJ: Educational Technology Publications.

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test and training efficiency. *Learning and Instruction*, *12*, 11–37.

Mory, E. H. (2003). Feedback research. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 745–783). New York: MacMillan Library Reference.

Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J. G., & Hummel, H. G. K. (2001). A model for optimizing step size of learning tasks in competency-based multimedia practicals. *Educational Technology Research and Development*, *49*, 87–103.

Nadolski, R. J., Kirschner, P. A., & van Merriënboer, J. J. G. (2005). Optimising the number of steps in learning tasks for complex skills. *British Journal of Educational Psychology*, *75*(2), 223–237.

Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J. G., & Wöretshofer, J. (2005). Development of an instrument for measuring the complexity of learning tasks. *Educational Research and Evaluation*, *11*(1), 1–27.

Naidu, S., & Bernard, R. M. (1992). Enhancing academic performance in distance education with concept mapping and inserted questions. *Distance Education*, *13*, 218–233.

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

Orlich, D. C., Harder, R. J., Callahan, R. C., Kauchak, D. P., & Gibson, H. W. (1994). *Teaching strategies: A guide to better instruction* (4th ed.). Toronto, Canada: Heath and Company.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology*, *86*, 122–133.

Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419–430.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, *18*, 16–25.

Quilicy, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *28*, 175–224.

Resnick, L. B. (1976). Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), *Cognition and instruction* (pp. 51–80). Hillsdale, NJ: Lawrence Erlbaum.

Romiszowski, A. J. (1981). *Designing instructional systems: Decision making in course planning and curriculum design*. London: Kogan Page.

Rowntree, D. (1992). *Exploring open and distance learning*. London: Kogan Page.

Salden, R., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. *Instructional Science*, *32*, 153–172.

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso, & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 113–150). New York: John Wiley & Sons.

Singley, M. K., & Anderson, J. R. (Eds.). (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith, P. L., & Ragan, T. J. (1999). *Instructional design* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1988). *Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains*. Tech. Rep. No. 441. Champaign, IL: University of Illinois. Center for the Study of Reading.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (1998). Instructional effects in complex learning: Do objective and subjective learning outcomes converge? *Learning and Instruction*, *8*, 117–129.

Sternberg, R. J., & Frensch, P. A. (Eds.). (1991). *Complex problem solving: Principles and mechanisms*. Hillsdale, NJ: Lawrence Erlbaum.

Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science*, *32*, 133–152.

Westera, W., & Sloep, P. B. (1998). The virtual company: Toward a self-directed, competence-based learning environment in distance education. *Educational Technology*, *38*, 32–37.

Wöretshofer, J., Nadolski, R. J., Starren-Weijenberg, A. M. A. G., Quanjel-Schreurs, R. A. M., Aretz, C. C. W. M., & van der Meer, N. H. W., et al. (2000). *Pleit voorbereid* [Preparing a plea] (Version 1.0) [multimedia CD-ROM]. *Heerlen*. The Netherlands: CIHO.