

The Effects of Practice Schedule and Critical Thinking Prompts on Learning and Transfer of a  
Complex Judgment Task

Anne Helsdingen<sup>a</sup>, Tamara van Gog<sup>b</sup>, and Jeroen van Merriënboer<sup>c</sup>

<sup>a</sup>Centre for Learning Sciences and Technologies, Open University of The Netherlands

<sup>b</sup>Institute of Psychology, Erasmus University Rotterdam, The Netherlands

<sup>c</sup>School of Health Professions Education, University of Maastricht, Maastricht, The Netherlands

Author Note:

Correspondence concerning this article should be addressed to Anne Helsdingen, Centre  
for Learning Sciences and Technologies, Open University of the Netherlands, P.O. Box 2960,  
6401 DL Heerlen, The Netherlands. T: +31 619348915, E: anne.helsdingen@ou.nl

Anne Helsdingen 1-10-10 00:25  
Met opmaak: Gecentreerd

### Abstract

Many instructional strategies that appear to improve learners' performance during training, may not realize adequate post test performance or transfer to a job. And the converse has been found true as well: instructional strategies that appear to slow the learner's progress during training often lead to better post-training or transfer performance. For example, many studies have shown beneficial effects of random over blocked practice on transfer of learning, even though blocked practice often leads to better performance during the training session. In a 2 x 3 factorial experiment ( $N = 120$ ) with the factors practice schedule (random, blocked) and critical thinking prompts (before task, after task, none), this study investigates whether this also applies to complex judgement tasks, and whether critical thinking prompts can enhance the effectiveness of particular practice schedules. It is hypothesized that prompts provided after task execution yield best transfer in a random practice schedule, whereas prompts provided before task execution yield best transfer in a blocked schedule. In line with our hypothesis, a blocked schedule led to better performance than random practice during training, but not on the transfer test, where a random schedule was beneficial. The hypothesized interaction effect was also found: critical thinking prompts after task execution significantly benefit transfer performance of participants following a random schedule, and transfer performance following a blocked schedule can be a little enhanced through providing critical thinking prompts before task execution. These results warrant instruction in critical thinking processes to teach complex judgment tasks, using random practice schedules combined with critical thinking prompts provided after task execution.

Keywords: contextual interference; critical thinking; complex judgment; learning; transfer.

## The Effects of Practice Schedule and Critical Thinking Prompts on Learning and Transfer of a Complex Judgment Task

A major aim of educational and training programs is to foster learning and transfer.

Learning is understood as the observable and enduring change in knowledge or behaviour as a result of experience (Skinner, 1950; Thorndike, 1910). This definition of learning incorporates both learning as a process of change and the outcome of that process: the durable change in knowledge or behaviour (Alexander, Schallert, & Reynolds, 2009). However, substantial changes in performance during training (apparent learning) are not always a good indicator of actual learning (Bjork, 1994; Bjork & Bjork, 2006). Post-training performance on test tasks that resemble the training tasks, and transfer to new tasks may be better indicators of a durable change in behaviour or knowledge. Transfer can be defined both in terms of the ability to flexibly apply (parts of) what has been learned to new tasks and/or new situations (see e.g., Detterman & Sternberg, 1993; Mayer, & Wittrock, 1996), as well as in terms of preparation for future learning (Bransford & Schwartz, 1999). Increasing interference between training tasks by sequencing different variations of the task randomly (i.e. contextual interference) may enhance post-training performance (Magill & Hall, 1990; Wulf & Shea, 2002). Research has also shown that instructional techniques to induce critical thinking (for a review, see Abrami et al., 2008) can foster learning and transfer. The present study investigates whether and how these techniques can be *combined* to foster learning and transfer of a complex judgment task.

### *Complex Judgment Tasks*

In the numerous choices or judgments people make every day, two distinct classes can be identified: Value judgments, which express their preferences, and predictions, which reflect what they expect to happen (Hogarth, 1980). Value judgments encompass for example a choice for

one house over another, or one pair of shoes over another. Predictions concern future outcomes, such as expectations regarding how someone might react to what you say or do, or who will win the next presidential elections. In this study we focus on predictive judgment tasks.

Consider for example an air defence officer who needs to judge the threat of an object, such as a hostile airplane. In making this judgment, several so-called 'points of reference' need to be considered that are expected to be related to estimating the threat of an object, such as its speed, position, altitude, direction, and identity. These points of reference are part of the officer's mental representation of the network of relationships between elements (objects, events) in the environment and the threat to be predicted. For example: How important is speed/identity/position? Is speed more important than identity when an object is near one's own position? What constitutes a higher threat, an object that is fast and far away but moving towards one's own position, or an object that is slow and very close but moving away? Accuracy of the judgment is assumed to depend on the extent to which the officer's mental representation matches the real network of relationships (Hogarth, 1980). It is this match, or lack thereof, that is the object of study within Social Judgment Theory (SJT; Brehmer & Joyce, 1988; Hammond, McClelland, & Mumpower, 1980), using policy capturing methods (for an overview see Dhami, Hertwig, & Hoffrage, 2004).

Within the SJT research paradigm, an experimental method was devised to study how people learn such difficult judgment tasks: The (Multiple) Cue Probability Learning experiment (MCPL; Björkman, 1965; Brehmer, 1972; Brehmer & Brehmer, 1988; Brunswik & Herma, 1951; Hammond, Hursch, & Todd, 1964; Hursch, Hammond, & Hursch, 1964; Smedslund, 1955), also referred to as Multidimensional Functional Learning (MFL; Hoffman, Earle, & Slovic, 1981). During a typical MFL experiment a person makes judgments based on a number

of probabilistic cues over a series of trials. Feedback may be given on each trial, or feedback may be given after subsets of trials. The aim is to learn to correctly predict the quantitative or categorical criterion value on each trial. Learning the relationships between one or more cues and the criterion requires that the learner compares different situations, with respect to the cue values and the criterion.

An important training goal for complex judgment tasks is not only learning the importance of different cues and their relationships, but also transfer of what has been learned into new tasks and future learning situations. Previous research has shown that both increasing contextual interference (Helsdingen, Van Gog & Van Merriënboer, 2010) and critical thinking instruction (Cohen, Freeman & Thompson, 1998; Freeman & Cohen, 1996; Helsdingen & Van den Bosch, et al., 2010) can enhance learning and transfer of complex judgment to novel tasks. We will discuss these techniques in more detail below.

#### *Contextual Interference*

Research has shown that some interventions that enhance performance during training may have detrimental effects on retention and transfer performance, and conversely, instructional manipulations that degrade performance during skill acquisition may support the long-term goals of training (for an overview, see Schmidt & Bjork, 1992). An example of the latter is by providing a practice schedule where different variations of the learning tasks are sequenced randomly (e.g., A-B-C-B-C-A-A-C-B) as opposed to sequenced in separate blocks (e.g., AAA-BBB-CCC) of related tasks (Shea & Morgan, 1979). This type of random sequencing is also often referred to as interleaving practice materials, or mixed practice (Hatala, Brooks, & Norman, 2003; Richland, Bjork, Finley, & Linn, 2005). The contextual interference effect refers to the consistent finding that increasing interference between training tasks by such random

practice may degrade performance during training but lead to better post-training performance and transfer than after blocked practice (Lee & Simon, 2003; Shea & Morgan, 1979; Wulf & Shea, 2002). It has mainly been studied in motor tasks (Brady, 1998; Cross, Schmitt, & Grafton, 2007; Guadagnoli & Lee, 2004; Lee & Magill, 1983; Magill & Hall, 1990; Shea & Morgan, 1979; Simon, 2007). However similar findings have been obtained with for example procedural tasks (Carlson, 1989; Carlson & Schneider, 1989; Carlson, Sullivan, & Schneider, 1989; Carlson & Yaure, 1990), cognitive operational tasks, such as interacting with automatic teller machines (Jamieson & Rogers, 2000), language learning (Jacoby, 1978), foreign vocabulary learning (Schneider, Healy, & Bourne, 1998, 2002), learning logical rules (Schneider, Healy, Ericsson, & Bourne, 1995), learning problem solving from worked examples (Paas & Van Merriënboer, 1994), and troubleshooting tasks (De Croock, Van Merriënboer, & Paas, 1998).

Several hypotheses have been advanced to explain the contextual interference effect (for an overview see e.g. Lee & Simon, 2004; Magill & Hall, 1990). The two most prominent explanations are the elaborative-processing hypothesis (Shea & Morgan, 1979) and the forgetting-and-reconstruction hypothesis (Lee & Magill, 1983). The elaborative processing hypothesis holds that random practice forces the learner into making intertask comparisons (Shea & Morgan; Shea & Titzer, 1993; Shea & Zimny, 1983, 1988; Lin, Fisher, Winstein, Wu, & Gordon, 2008). The need to identify each task variation that is presented causes the disadvantage during learning, whereas in blocked practice, the continued repetition of long series of the same task variation makes it less important to keep track of which task is which. Thus, blocked practice is less demanding than random practice, and as a consequence, blocked schedules often yield better performance during practice than random schedules. However, the need to compare and contrast in a random schedule, leading to more elaborate and distinctive memorial

representations of the practiced tasks yields superior performance in retention tests. In transfer tests, the argument is that random practice has made learners more adept at identifying the relevant features of the to-be-performed transfer task, providing an advantage on these novel tasks.

The forgetting–reconstruction hypothesis (Lee and Magill, 1983, 1985) states that random practice forces learners to discard the task strategy between tasks and either retrieve or reconstruct a new strategy for successive tasks. In blocked practice, a given strategy can be planned and maintained in working memory across an entire block of trials. Although modifications may be made to the strategy, this practice schedule generally affords the learner only one opportunity to bring up, or construct, each strategy. The uninterrupted repetitions of the a similar strategy in blocked practice makes for relatively high-quality performance, but the lack of practice at constructing the strategies anew supports relatively poor learning. In random practice, the opposite is seen: the need to continually reconstruct the to-be-performed action pattern from one trial to the next makes for lower quality performance, but affords an advantage in delayed tests of learning, which make high demands on such reconstruction abilities.

Both the elaborative and the reconstruction processing are assumed to take place in working memory during the intertask intervals (Lee & Simon, 2004). A recent study by Lin, Fisher, Winstein, Wu, & Gordon, (2008) has shown that when working memory processes are disrupted (with transcranial magnetic stimulation) during intertask intervals, the learning benefit of random practice disappeared, whereas the learning value of a blocked practice schedule was not enhanced by the intertask disruptions. These results underpin the elaborative processing hypothesis and fail to support the forgetting-reconstruction hypothesis.

Another instructional intervention that may encourage learners to consider more task features, compare between tasks and thus enhance learning and transfer test performance is critical thinking instruction.

### *Critical Thinking Instruction*

Critical thinking is conceptualized as higher order thinking that is purposeful, reasoned, and goal directed. It entails the ability to recognize the fallibility of one's own opinions (Nickerson, 1987), to reason from starting points with which one disagrees (Norris & Ennis, 1989) or to accept statements as true even when they don't agree with one's own position (Zechmeister & Johnson, 1992). It is involved in solving problems, formulating inferences, calculating likelihoods, and making decisions (Frijters, Ten Dam, & Rijlaarsdam, 2008; Halpern, 2003). Within the critical thinking research community, there is an ongoing debate about the issue of generalizability or specificity of critical thinking skills (Kuhn, 1999; Siegel, 1991). Some studies have shown that special critical thinking programs do not usually result in long lasting effects (Perkins & Salomon, 1989; Singley & Anderson, 1989; Tsui, 1999). But at the same time, several authors point out that some general principles of critical thinking transcend specific subjects (e.g. Klaczynski, 2001; Stanovich & West, 2000; Tsui, 1999). Transfer of skills across different task and knowledge domains is important because today's jobs require professionals to work in continuously changing contexts. Therefore, the challenge is to teach critical thinking in the context of specific meaningful subject matter, but in such a way that transfer to other domains becomes possible (Brown, 1997; Frijters et al., 2008; Halpern, 1998).

There are many different types of critical thinking instruction (see e.g., Abrami et al., 2008; Facione, 1990), ranging from general approaches in which critical thinking strategies are explicitly and domain-independently taught (Ennis, 1989) to very domain specific courses on a

specific subject in which the general critical thinking processes are not made explicit (Abrami, et al., 2008). The method of critical thinking instruction developed by Cohen and Freeman (1998) explicitly teaches general critical thinking skills using domain specific problems. It focuses on recognition-based decision making as well as (meta)cognitive techniques that guide and critically review this decision-making process (Cohen, Freeman, & Wolf, 1996).

People's decisions are based on recognition of aspects of a situation, matching recognized aspects with earlier experiences (Kuhn, 1999), and forming a mental model of the current situation that implies a prototypical or sufficient decision option (Lipshitz & Ben Shaul, 1997). However, this mental model will often be incomplete or inconsistent because of missing information, conflicting evidence, and unreliable assumptions. Critically testing and evaluating one's mental model are therefore considered paramount in the decision-making process, especially when high stakes are involved, when problems are dynamic and complex, or both. Judgment based on recognition of a situation requires exposure to many different problems during training. Furthermore, training should foster the (meta)cognitive skills that allow decision makers to verify and improve the results of recognition-primed decision making processes, by identifying evidence-conclusion relationships (i.e., arguments), critiquing the arguments used to support a conclusion (i.e., identifying incompleteness, conflict, or unreliability), and correcting the errors in argumentation (Cohen et al., 1996).

The method for critical thinking instruction described below typically involves an explanation of the decision making process beforehand and an instructor who prompts the learner to reflect on his or her cognitive strategy and initiate one of the critical thinking steps during learning. These steps include (Freeman & Cohen, 1996; see also Helsdingen, Van den Bosch et al., 2010):

*Creating a story:* A story is a comprehensive assessment of the situation, in which all existing evidence is incorporated and explained and assumptions are made about uncertain aspects of the situation. Past, present and future are addressed in the story. The purpose of story building is to keep participants from assessing situations solely on isolated events. Instead, participants are taught how they can integrate the available information into its context.

*Testing a story:* Participants are instructed how to identify inconsistency and uncertainty, and how to adjust or refine their story by deliberate testing. They have to correct these problems by collecting more data, retrieving knowledge from memory, making assumptions about the missing piece of the story, or by resolving conflicts in the argumentation.

*Evaluating a story:* After a story is constructed and tested for gaps, it should be evaluated for its plausibility. The decision maker has to take a step back, identify critical or hidden assumptions and play the devil's advocate by falsifying these assumptions, that is, explaining how an assumption can be false and build an alternative story.

*Quick test:* Critical thinking is not always appropriate. Decision makers have to evaluate the time available and the consequences of their actions. In stressful situations such as those often encountered by professional decision makers, there is usually little time to spare. The decision makers should act immediately *unless* the risk of delay is acceptable, the costs of errors are high (i.e., when lives are at stake), and/or the situation is non-routine or problematic (i.e., complex or novel problems).

Instructors usually deliver prompts when they think they are required during the training. Such (meta)cognitive prompting in itself can be effective for learning: In the educational literature, positive effects are mentioned of, for example, planning prompts (Davis, 2003), driving questions (Dochy, 1992; Douglas, Hosokawa, & Lawler, 1988; Morgan & Saxton, 1991;

Naidu & Bernard, 1992; Orlich, Harder, Callahan, Kauchak, & Gibson, 1994; Rowntree, 1992), deep level reasoning questions (Craig, Sullins, Witherspoon, & Gholson, 2006), and reflection prompts (e.g., Belland, Glazewski, & Richardson, 2007; Davis & Linn, 2000; Van den Boom, Paas, & Van Merriënboer, 2007; Van den Boom, Paas, Van Merriënboer & Van Gog, 2004).

Cohen and colleagues (Cohen et al., 1998; Freeman & Cohen, 1996) attributed the beneficial effects of critical thinking instruction on learning to participants having acquired an appropriate decision making strategy. When students are taught critical thinking skills, they are less likely to make the typical judgment mistakes that are caused by for example confirmation bias or failures to reason adequately about probabilities (Halpern, 1997). Thus, they will show better overall performance on judgment and decision-making. However, in a study by Helsdingen, Van den Bosch et al., (2010) in which naval warfare officers learn to identify, categorize, and react to naval surface threats in a dynamic and probabilistic simulated environment, critical thinking instruction was found to have effects primarily on transfer performance; the test included real-world events presented in a simulated environment. On this transfer test, officers who received critical thinking instruction were better at selecting relevant information, identifying missing or conflicting evidence, providing arguments for their assessments, planning their actions in time, and displayed better team skills than officers who received the same amount of judgment training but without critical thinking instruction. This may be an indication that apart from having acquired a better decision making strategy, participants also gained a more abstract level of understanding of the task content that enabled them to solve decision problems that differed from the learning problems on a superficial and a structural level.

Critical thinking instruction that includes (meta)cognitive prompts may have beneficial effects on learning and transfer because it might facilitate processes such as self explanation and reflection, which have been proven to enhance learning and transfer (Alevan & Koedinger, 2002; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; Ferguson-Hessler & De Jong, 1990; Lee & Hutchinson, 1998; Renkl, 1997). Findings by Schworm and Renkl (2007) seem particularly relevant for complex judgment tasks: they demonstrated that self-explanation is a promising method of enhancing skills in ill-structured domains requiring argumentation. Chi et al. (1989) attributed the positive effects of self-explanation during training tasks on learning and transfer performance to learners being able to judge their learning progress adequately, and thus be able to identify their additional learning needs to attain their ultimate learning goal. More recently, Chi (2000) has proposed that self-explanations may also play an important role in testing and repairing one's own mental model. Although this explanation was based primarily on text comprehension research, it may also play a role in the effectiveness of the critical thinking method developed by Cohen et al. (1996). They refer to creating, testing, and evaluating a 'story', but this 'story' reflects a mental model of a situation.

#### *The Present Study*

As mentioned before, previous research has shown that both increasing contextual interference (Helsdingen, Van Gog et al., 2010) and critical thinking instruction (Cohen et al., 1998; Freeman & Cohen, 1996; Helsdingen, Van den Bosch et al., 2010) can enhance learning and transfer of complex judgment to novel tasks. The present study investigates whether and how these instructional techniques can be combined to foster learning and transfer of complex

judgment, not only in terms of transfer to novel tasks but also in terms of preparation for future learning.

It does so by examining the effects of random and blocked practice schedule in combination with no critical thinking instruction or critical thinking instruction involving either proactive prompts (i.e., given before execution of several tasks) or retrospective prompts (i.e., given after execution of several tasks) next to a general instruction, using an MFL-task in the domain of criminal investigation. Participants' task is to read a case description and judge its priority for the police on the basis of several cues that appear in each case description. Participants received feedback about the actual priority after they gave their estimate; thus, by comparing the different cue values and priorities of subsequent cases, participants could learn to identify the relevant cues and how these related to the priority. In a blocked practice schedule, only one cue changed value from one case to the next, thus providing the participants the opportunity to see how this particular cue changed the priority of a case. In a random practice schedule, none, one, or several cues changed value from one case to the next, requiring participants to focus on all cues with each presentation of a case and compare their values to the previous presentation. Hence, participants have to process more information in the random practice schedule, and this may result in more elaborate representations of the domain.

Critical thinking prompts draw participants' attention to the similarities and differences between consecutive cases and how these differences affect the priority. It is hypothesized that in order to be effective in combination with different practice schedules, timing of the critical thinking prompts is crucial. In a blocked practice schedule, critical thinking processes would need to be prompted *before* execution of the task (i.e., proactive), so that critical thinking may induce the intertask elaborative processes that are normally absent in blocked practice (Lin,

Fisher, Winstein, Wu, & Gordon, 2008). In a random practice schedule, the key factor in successful planning of critical thinking prompts is the learners' limited processing capacity: When learners already enjoy the learning benefit of the intertask elaborative processes that are induced by the random schedule, few if any cognitive resources are available for critical thinking processes (cf. Sweller, Van Merriënboer, & Paas, 1998). However, prompting critical thinking after execution of the training task (i.e., retrospective) may have additional beneficial effects on learning and transfer without leading to processing capacity problems.

In sum, it is hypothesized that *during* learning a blocked practice schedule generates better performance than a random practice schedule. Regarding learning *outcomes*, measured in terms of post-test and transfer test performance, it is hypothesized that in the conditions without critical thinking instruction and prompts, random practice will be more effective than blocked practice. In the conditions with critical thinking instruction and prompts, it is expected that learners in a blocked practice schedule profit more from proactive prompts than from retrospective or no prompts, while in a random practice schedule, learners benefit more from retrospective prompts than from proactive or no prompts.

## Method

### *Participants and Design*

Participants were 120 students from different departments of two large Dutch universities (63 males and 57 females). Their age varied from 20 to 25 years ( $M = 21.58$ ,  $SD = 1.59$ ). Participation was voluntary and rewarded with 45 Euros. Participants were randomly assigned to one of the six conditions in a 2 x 3 factorial design with the factors Practice Schedule (Blocked or Random) and Critical-Thinking Prompts (Proactive, Retrospective, or None): (a) blocked practice schedule with proactive critical-thinking prompts ( $n = 20$ ), (b) blocked practice schedule

with retrospective critical-thinking prompts ( $n = 20$ ), (c) blocked practice schedule without prompts ( $n = 19$ ), (d) random practice schedule with proactive critical-thinking prompts ( $n = 20$ ), (e) random practice schedule with retrospective critical-thinking prompts ( $n = 21$ ), and (f) random practice schedule without prompts ( $n = 20$ ).

### *Materials*

*Training tasks.* A set of 32 case descriptions of crimes was developed for this experiment. Each case had to be prioritized on the urgency for the police to deal with it. Priorities could be determined on the basis of the dichotomous values on three different cues that occurred in each case description: (a) the condition of the victim (injured, dead), (b) the nature of the crime (burglary, violence/holdup), and (c) the use of a weapon (no firearm, firearm). The upper part of Table 1 presents the priority scores for each combination of cue values. Four crime descriptions were developed for each of the eight combinations of cue values, resulting in 32 cases (see Figure 1 for an example of a case). Case descriptions, relevant cues, and their effect on priorities were evaluated and approved by a police officer on being representative for the domain<sup>1</sup>.

The cases were presented one by one on a computer screen (see Figure 1). On the lower half of the screen, beneath the presentation of the case, a slide bar with an indicator of priority was presented. The slide bar covered the whole range of priority scores: 1-100. There were no numbers or other reference points on or near the slide bar. The participants could manipulate an indicator on the slide bar with the computer mouse to mark the priority of a case. In addition,

---

<sup>1</sup> The relationships between cues and priorities of cases were artificial: The model underlying the relationships was constructed specifically for experimental purposes. However, since it was important that the task was representative of real-world judgment tasks, the model resembled real-world relationships in content, importance and direction of the relationships.

they could adjust the margin of the area on the right and the left from the indicator (minimum = 2, maximum = 50).

Feedback during training consisted of a second slide bar with the indicator at the position of the true priority score. This feedback slide bar was presented above the first one, directly after participants indicated their priority and margins. If the true priority score was within the margin of the participant's response, a green rectangle was presented around the feedback slide bar. If the true priority score lay outside this margin, a red rectangle was presented around the feedback slide bar.

*Practice schedules.* In the blocked practice schedule, the order of 32 training tasks was the same for all participants. The 32 training tasks were sequenced in such a way that in the first 12 tasks only one cue changed its value from one task to the next, and the cue that changed its value was alternated from task to task. Thus, if the first task involved a dead victim (cue 1) after a burglary (cue 2) and the use of a firearm (cue 3), the second task could be a dead victim (cue 1: unchanged) after a burglary (cue 2: unchanged) *without* the use of a firearm (cue 3: changed), and the third task could subsequently be an *injured* victim (cue 1: changed) after a burglary (cue 2: unchanged) without the use of a firearm (cue 3: unchanged). In the following 12 tasks two cues simultaneously changed value from one task to the next. In the last 8 tasks all three cues changed value from one case to the next. In the random practice schedule the training tasks were presented in a randomized order that was the same for all participants in this schedule.

*Critical-thinking instruction and prompts.* Participants in all critical thinking conditions were introduced to critical thinking with a paper-based instruction (see appendix A) explaining four steps: (a) creating a model of the relationships between the cues and the priorities, (b) testing a model, (c) evaluating a model, and (d) finding alternative explanations (i.e.,

contingencies). Apart from this textual explanation, a handout with a graphic representation of the four steps (Figure 2) and two short stories of critical thinking in real-world situations were provided.

Both the 12 proactive and the 12 retrospective prompts were formulated as open-ended questions provided on the computer screen, which intended to focus the participant's attention to particular cues and how their priority changed (for examples, see Table 2). The proactive prompts preceded every two or three cases, similar retrospective prompts followed those same cases, thus, only the timing relative to the training tasks was different for both critical thinking conditions. Participants were told that the open-ended questions could help them learn the task and they were therefore required to answer the questions for themselves.

*Post-test and transfer test.* The post-test consisted of 8 tasks that were similar to the training tasks in structural features (same combination of cues and same domain, i.e., crimes) but different to the training tasks in surface features (i.e., cover stories). The post-test tasks were presented in the same random order to all participants and the sequence contained no critical-thinking prompts and no feedback.

The transfer test consisted of 32 tasks describing traffic offenses that differed in both structural and surface features from the training tasks (see appendix b or <http://files.me.com/anne.h/a0thyu>). Priorities depended on: (a) nature of the offence (speeding, driving without insurance), (b) condition of the offender (drunk, not drunk), and (c) history (first offence, recidivist). These transfer tasks were scheduled in random order that was the same for all participants and no critical-thinking prompts were given. However, feedback was provided because the new domain contained new cues to base the priority on. See the bottom part of Table 1 for an overview of the cues and priorities of the training, post-test, and transfer tasks.

### *Procedure*

The experiment had a duration of approximately two hours and was run in sessions with at most six participants. Before the experiment started, participants read a short instruction explaining how to rate the priority of the cases by means of the slide bar and how to indicate the margin of certainty. All participants received paper and pencil to make notes. In addition, the critical-thinking conditions received a written instruction about the four steps of the critical-thinking method, the handout with the graphical representation of the four steps of critical thinking, and two example stories of critical thinking in real world situations. Then, all participants were assigned to a computer and started to work on the training cases of their respective condition. The post-test was administered right after completion of the training tasks, and the transfer test immediately followed the post-test. Participants could choose their own pace; the time they took to complete the whole experiment was logged. For both training tasks and test tasks, participants' *deviation score*, defined as the absolute difference between the estimated priority and the true priority of the case, was automatically stored in a log file. After completion of the transfer test open exit interviews were conducted with the participants.

### Results

For training, post-test and transfer, the mean deviation scores for each *phase of 8 consecutive cases* were calculated for each participant. Time-on-task spent on the entire experimental session did not differ between conditions,  $F(1, 114) = .01, p > .20$ . In all analyses reported below, a significance level of .05 is set, and partial eta-squared or Cohen's  $d$  are reported as a measure of effect size.

### *Training*

Participants' deviation scores over all conditions ranged from 11.90 to 31.06 ( $M = 19.63$ ,  $SD = 3.92$ ). Table 3 presents the means and standard deviations of the deviation scores per phase (of 8 cases) and condition. A repeated measures ANOVA with Practice Schedule and Critical Thinking Prompts as between-subjects factors and Phase (four consecutive phases of 8 tasks) as within-subjects factor was conducted to analyze the effects on participants' deviation scores over time during training. There was a substantial main effect of Phase,  $\lambda = .217$ ,  $F(3, 112) = 135.046$ ,  $p < .01$ ,  $\eta_p^2 = .783$ , with all groups showing a reduction in deviation scores (i.e., performance improvement) across the four phases. There was also a main effect of Practice Schedule,  $F(1, 114) = 14.431$ ,  $MSE = 52.998$ ,  $p < 0.01$ ,  $\eta_p^2 = .112$ . As predicted, participants in the blocked practice schedule had lower deviation scores (better performance) than participants in the random practice schedule. There was no main effect of Critical Thinking Prompts,  $F(2, 114) = 2.225$  *ns*.

There was no significant three-way interaction of Phase, Practice Schedule and Critical Thinking Prompts,  $\lambda = .919$ ,  $F(6, 224) = 1.614$ , *ns*. There were, however, significant interactions between Practice Schedule and Phase,  $\lambda = .847$ ,  $F(3, 112) = 6.732$ ,  $p < .01$ ,  $\eta_p^2 = .153$ , and between Critical Thinking Prompts and Phase,  $\lambda = .813$ ,  $F(6, 224) = 4.060$ ,  $p < .01$ ,  $\eta_p^2 = .098$ , indicating that performance development over time was different for the different practice schedules and for the different critical thinking instruction conditions. Post hoc *t*-tests for the effects of Practice Schedule on performance showed that participants in the blocked practice schedule ( $M = 18.53$ ,  $SD = 6.42$  and  $M = 16.08$ ,  $SD = 5.89$ ) outperformed participants in the random practice schedule ( $M = 24.34$ ,  $SD = 5.71$  and  $M = 18.34$ ,  $SD = 4.59$ ) only in phases 2 ( $t(118) = -5.24$ ,  $p < .01$ ,  $d = 0.96$ ) and 3 ( $t(118) = 2.34$ ,  $p = .02$ ,  $d = 0.42$ ). Post-hoc tests for the effects of Critical Thinking Prompts on performance per phase (of 8 consecutive cases) showed

for the 12 comparisons only one significant difference: in phase 2, participants who received retrospective prompts ( $M= 19.39, SD= 1.03$ ) showed significantly higher performance ( $p = .04, d = 3.54$ ) than participants who did not receive prompts ( $M = 23.07, SD = 1.10$ ), other conditions did not differ from each other (see Table 4).

There was no significant interaction between Critical Thinking Prompts and Practice Schedule,  $F(2, 114) = 2.643, ns$ .

#### *Post-Test*

Participants' deviation scores over all conditions ranged from 3.79 to 29.29 ( $M = 14.99, SD = 5.58$ ). Table 5 presents the means and standard deviations of the deviation scores on the post-test. A 2 x 3 ANOVA on the mean deviation scores showed that Critical Thinking Prompts had no main effect on performance on the post-test,  $F(2, 114) = 2.59, ns$ . There was a main effect of Practice Schedule,  $F(1, 114) = 4.23, MSE = 120.39, p = .04, \eta_p^2 = .04$ : Participants that followed a random practice schedule reached lower deviation scores ( $M = 14.04, SD = 0.68$ ) on the post-test (i.e., better performance) than participants that followed the blocked practice schedule ( $M = 16.04, SD = 0.70$ ). However, there was also a significant interaction between Practice Schedule and Critical Thinking Prompts,  $F(2, 114) = 3.26, MSE = 92.78, p = .04, \eta_p^2 = .05$ . This interaction was explored by analysis of both practice schedules separately. Analysis of the random practice schedule showed a significant effect of Critical Thinking Prompts,  $F(2, 58) = 5.98, MSE = 153.70, p < .01, \eta_p^2 = .17$ , with retrospective prompts ( $M = 10.97, SD = 5.81$ ) yielding significantly better ( $p < .01, d = 1.01$  and  $p = .05, d = 0.74$ ) results than both proactive prompts ( $M = 16.21, SD = 4.45$ ) and no prompts ( $M = 14.94, SD = 4.81$ ). In the blocked practice schedule, Critical Thinking Prompts had no differential effects on performance,  $F(2, 56) = .48, ns$ .

### *Transfer Test*

Participants' deviation scores over all conditions ranged from 7.90 to 34.00 ( $M = 18.08$ ,  $SD = 6.10$ ). Table 6 presents the means and standard deviations of the deviation scores during the four consecutive phases of 8 tasks of the transfer test. A repeated measures ANOVA with Practice Schedule and Critical Thinking Prompts as between-subjects factors and Phase (four consecutive phases of 8 tasks) as within-subjects factor was conducted to analyze the effects on participants' deviation scores over time during the transfer test. There was a substantial main effect of Phase,  $F(3, 112) = 36.851$ ,  $p < .01$ ,  $\eta_p^2 = .497$ , with all groups showing a reduction in deviation scores (i.e., performance improvement) over consecutive phases. There was also a main effect of Practice Schedule,  $F(1, 114) = 38.203$ ,  $MSE = 102.350$ ,  $p < 0.01$ ,  $\eta_p^2 = .251$ . As predicted, participants in the random practice schedule had lower ( $M = 15.29$ ,  $SD = 5.59$ ) deviation scores (i.e., better performance) than participants in the blocked practice schedule ( $M = 20.96$ ,  $SD = 5.24$ ). There was no significant main effect of Critical Thinking Prompts,  $F(2, 114) = 1.839$  *ns*.

There was a significant interaction between Practice Schedule and Critical Thinking Prompts,  $F(2, 114) = 8.972$ ,  $MSE = 102.350$ ,  $p < 0.01$ ,  $\eta_p^2 = .136$ . Further investigation of this interaction by analyzing random and blocked practice schedules separately, showed that Critical Thinking Prompts significantly affected participants' transfer score, both in a random practice schedule,  $F(2, 58) = 4.13$ ,  $MSE = 116.73$ ,  $p = .02$ ,  $\eta_p^2 = .12$ , and a blocked practice schedule,  $F(2, 56) = 6.94$ ,  $MSE = 158.18$ ,  $p < 0.01$ ,  $\eta_p^2 = .20$ . In line with our hypothesis, in a random practice schedule, retrospective critical thinking prompts ( $M = 12.49$ ,  $SD = 4.95$ ) yielded better scores (resp.  $p = .04$ ,  $d = 0.78$ , and  $p = .04$ ,  $d = 0.82$ ) on the transfer test than both proactive prompts ( $M = 16.64$ ,  $SD = 5.66$ ) and no prompts ( $M = 16.68$ ,  $SD = 5.29$ ), whereas in a blocked practice

schedule, proactive prompts ( $M = 17.80$ ,  $SD = 3.03$ ) lead to better scores (resp.  $p < .01$ ,  $d = 1.21$ , and  $p < .02$ ,  $d = 0.95$ ) than both retrospective prompts ( $M = 23.16$ ,  $SD = 5.48$ ) and no prompts ( $M = 21.98$ ,  $SD = 5.44$ ).

Other interactions between Phase, Practice Schedule, and Critical Thinking Prompts,  $A = .982$ ,  $F(6, 224) = .347$ , *ns*, Phase and Critical Thinking Prompts,  $A = .959$ ,  $F(6, 224) = .795$ , *ns*, or Phase and Practice Schedule,  $A = .986$ ,  $F(3, 112) = .544$ , *ns*, were not significant.

To investigate whether the effects of Practice Schedule and Critical Thinking are present from the beginning of the transfer test, or occur only after a few tasks are executed, we examined participants' performance during the first Phase of the transfer test in more detail. Figure 3 presents the performance data of participants in the 6 different conditions on the first 8 tasks. From Figure 3 it becomes clear that all participants start off performing at approximately the same level, and the effects of Practice Schedule and Critical Thinking Instruction appear after 2 tasks. Figure 3 and the pairwise comparisons between the 6 groups (Table 7) clearly show 3 subgroups in the performance development in that first phase: The blocked practice group without prompts and with retrospective prompts are slowest in their performance development, the random practice with retrospective prompts learned the new traffic cases quickest, and the groups with random practice without prompts, or with proactive prompts are between the slowest and quickest developing groups.

### Discussion

This study examined the effects of critical thinking instruction involving proactive and retrospective critical thinking prompts provided in combination with random and blocked practice schedules on learning a multidimensional functional training task. With respect to participants' performance it was hypothesized that: (a) A blocked practice schedule yields better

performance during training than a random practice schedule and (b) a random practice schedule results in best post-test and transfer performance in conditions where no critical thinking instruction is provided. However, when such instruction is provided, it was hypothesized that (c) blocked practice schedule yields higher transfer when participants receive proactive critical thinking prompts rather than retrospective or no prompts, but (d) a random practice schedule yields higher transfer when participants receive retrospective prompts rather than proactive or no prompts.

In line with our first hypothesis, results showed that in training, participants who followed a blocked practice schedule performed better than participants who practiced according to a random schedule. This is consistent with research findings on the effects of contextual interference in other domains such as motor skills (Brady, 1998; Magill & Hall, 1990), troubleshooting skills (De Croock, Van Merriënboer, & Paas, 1998; Jelsma, Van Merriënboer, & Bijlstra, 1988), or language skills (Schneider, Healy, & Bourne, 2002). There was an interaction between Phase (i.e. the consecutive phases of 8 cases each) and Practice Schedule, subsequent post-hoc tests showed that the difference in performance between blocked and random practice schedule was not present at the start of training: this indicates that the groups started off at the same baseline performance. And at the end of training, again no difference was found between performance of participants in the blocked and random practice schedules. This can be explained by the fact that the last training tasks in the blocked schedule were scheduled in such a way that all cues changed value from one task to the next. Apparently, learners in this blocked group were not very adequate in dealing with these changes.

Looking at results on the post-test and transfer test, we see that participants who followed a random practice schedule showed better performance. This is in line with our second

hypothesis. This means that the random practice schedule not only prepared participants for the random schedule of the post-test tasks, but also improved their ‘future learning’ on the traffic cases of the transfer test. Evidently, participants who had followed a random practice schedule are able to identify relevant cues and estimate their impact on the priority of a case quicker on a subsequent transfer task than participants that had followed a blocked schedule. When we examined the first 8 transfer tasks in a little more detail, we found that there was not a first trial difference between the practice schedule conditions, but differences in performance became evident after two tasks. This makes sense, because to be able to identify relevant cues from the irrelevant ones in the case descriptions requires at least a few presentations to see which cues recur in the next cases. We assume that participants from the blocked practice group were not adequately trained in how to approach the new cases because their practice schedule did not require them to consider many cues at the same time and compare their values between subsequent tasks, thus leading to less elaborate mental representations of how to approach cue-criterion tasks. Random practice, on the other hand, forced participants to search for (in-)consistencies between subsequent tasks and relating that to the criterion. This provided them with more sophisticated schema on cue-criterion tasks and how to approach them.

The results on the post- and transfer test also showed that participants in the random practice schedule benefited from retrospective critical thinking prompts, but not from proactive prompts. It seems that when participants experienced a high load from the random practice schedule, they were not able to concurrently apply the critical thinking steps. In fact, the results show no difference between the no prompts and the proactive prompts condition for a random schedule, which indicates in the random practice, participants were able to completely ignore the proactive prompts. This reflects other research findings (e.g., De Bruin, Rikers & Schmidt, 2005;

Kanfer & Ackerman, 1989; Schnotz & Kürschner, 2007; Winne, 1995), showing that specific learning strategies and processes such as self-monitoring and self-regulation will only benefit learning when learners have considerable resources available for such metacognitive activities. Similarly, critical thinking can only be practiced in training situations where participants have sufficient resources available. This may partly explain why the results did not show a main effect of Critical Thinking Prompts: When participants were overwhelmed by task demands, they ignored prompts that provided additional load, thereby creating a situation similar to the No Prompts condition. Blocked practice, on the other hand, was easy, and, as Schnotz and Kürschner, (2007) concluded, for such easy tasks there is no need for “search for patterns in the learning material, restructuring of problem representations, or metacognitive processes”. Thus, as becomes apparent from our the transfer test results, unless specifically prompted to engage in critical thinking *before* the execution of a task (proactively), participants were reluctant or unable to do so *after* task execution (retrospectively) causing the Retrospective Prompts condition to resemble the No Prompts condition.

However, contrary to performance on the transfer test, post-test performance was not affected by critical thinking prompts. The explanation for this finding probably lies in the nature of the post-test tasks. Because these tasks did not differ from the training tasks on structural features, participants in all blocked conditions could rely on their specific memories of similar cases they had experienced during training. Summarizing, Critical Thinking Prompts mainly benefit transfer performance when these prompts are appropriately timed in a specific practice schedule, such that overload is avoided in demanding task schedules and active engagement in critical thinking is pushed in easy task schedules.

A limitation of this study is that we have only an overall time-on-task measure, so it is possible that there may have been differences in the amount of time needed to learn the transfer test cases between conditions. In that case, however, these time differences must have been compensated by differences in the opposite direction in the tests because there were no significant differences in overall time-on-task.

The finding that relative novices were able to profit from the critical thinking instruction and prompts while learning a complex judgment task, even in a random practice schedule which already increases cognitive load (Paas & Van Merriënboer, 1994), is remarkable. Interestingly, their performance was not *negatively* affected from proactive prompts during a random schedule, which suggests that in high workload conditions learners may be able to ignore given prompts all together and focus on learning the specific rules and principles of the task without being distracted. However, this is speculation, and a limitation of this study is that the responses to critical thinking prompts were not recorded either by having participants respond to the prompts in writing (cf. Schworm & Renkl, 2007) or by means of verbal protocols (cf. Ericson & Simon, 1993; Van Gog, Paas, Van Merriënboer, & Witte, 2005). Such data might have provided more direct evidence for our assumptions as to why retrospective prompting is effective with random practice and proactive prompting is effective with blocked practice.

Within the critical thinking research community there is ongoing discussion about whether explicit instruction is required to enhance critical thinking skills or whether mere immersion is effective (see e.g., Abrami et al., 2008), and on whether this skill should be taught in a domain-general way or within a specific task or domain (see also e.g., Gagné, 1984). The results of our study provide support for the benefit of explicit instruction within a specific task or domain. This is in line with results of a recent study by Angeli & Valanides (2008). In their

study, Angelis and Valanides compared a 1) general content free method with 2) an approach in which critical thinking methods were taught within domain specific problems, 3) a method in which domain specific problems were given without explicit teaching of critical thinking strategies and 4) a control group. Their results showed superior performance of deliberate instruction targeting the teaching of critical-thinking skills embedded within a domain specific problem. Teaching of critical thinking skills in a de-contextualized way and with no support or feedback from the researcher, did not seem to result in improvement of learner's critical thinking skills (Angeli & Valanides, 2008).

The results of our study suggest that explicit instruction in critical thinking is effective, even without feedback on the critical thinking strategies of learners and without the social context, e.g. group discussion or dialogue between learners and an instructor, as is part of the collaborative reasoning approach to critical thinking of Clark et al. (2003). However, based on this study we cannot disentangle whether the critical thinking skills themselves transferred, or whether critical thinking instruction lead to deeper, more principled understanding of the how to approach the type of cue- criterion problems, that is, how to identify which cues are relevant and how to compare between cases in order to assess the relative weight of cues.

Nonetheless, the findings reported here are highly relevant for educational and training settings. Training programs for professional decision makers, such as for example military commanders, involve scenario based exercises that are preceded by theory classes in which the rules of the domain are explained (Knutzon, Walter, Sannier, & Oliver, 2004). The scenario-based exercises are field or simulator exercises in which learners receive a role and assignment, and problems are introduced by events that occur during the course of the scenario. The preceding theory classes only tackle the simple or unambiguous rules (e.g. the Rules of

Engagement, military doctrine, or simple if-then rules); whereas in the theatre of war, the complexities arise from the interdependencies between these rules and from many more complex and ambiguous rules that are not explicitly taught but rather have to be learnt by practice with many different scenario based exercises (Drillings & Serfaty, 1997). Therefore, our study has taken the scenario-based exercise as a point of reference. The interventions, studied here, however, may also have beneficial effects with other types of instruction that involve a higher degree of guidance than our learning by doing approach, such as worked examples. As Paas and Van Merriënboer (1994) have demonstrated, instruction with worked examples can be enhanced by introducing variability of practice. Also, prompting self-explanations has been shown to benefit instruction from worked examples (Schworm & Renkl, 2007), but these interventions have not been studied in combination. An interesting question for future research would be whether the effects of combining contextual interference and critical thinking would be similar in a replication of the present study with worked examples, because these are known to lower cognitive load compared to problem-solving (Paas, 1992; Van Gog, Paas, & Van Merriënboer, 2006) and therefore, outcomes regarding for example the combination of random practice and proactive prompts might differ.

However, in considering the practical implications of our results, we have to take into account our operationalization of a blocked practice schedule, which is somewhat different from blocked schedules as they are traditionally designed in contextual interference studies. In a blocked schedule, the sequence of training tasks is such that only one task variation per phase (in our study 8 cases) is practiced (e.g. block 1: AAAAAA, block 2: BBBBBB, where A and B are variations of the training task). As we stated in the Introduction, in complex judgment, the task for the judge is to predict a future outcome on the basis of a few cues. In learning to predict such

future outcomes, the judge has to learn which cues are relevant (learn cue selection), how each cue relates to the criterion to be predicted, and whether cues are intercorrelated (i.e., learn cue-criterion functions). Thus, the creation of blocks for these complex judgment tasks is less straightforward than it would be for less complex tasks. The blocked schedules in our experiments, were such that in the first phases one cue changed value from one task to the next. In terms of predicting the future outcomes, this would not seem to be a blocked sequence as the resulting sequential tasks are different task variations. However, in terms of the training task, that is, to learn cue-criterion functions, participants are provided the opportunity to learn the effects of one cue-criterion function before moving on to the next. And in the last phase, two (or more) cues changed value from one case to the next, providing the opportunity to learn the interaction between the cues. The blocked schedule may encourage participants from the start to adopt a strategy to focus on the cue that has changed with each new case. Whereas after being trained with a random schedule, participants are looking for all the differences between the former cases and the current case.

Although our operationalization of blocked practice differs from other designs in contextual interference studies, it does resemble real world training approaches for complex judgment and decision-making. For example, sonar image operators, in their training to identify and judge sonar contacts, first learn how a sonar image depends on ocean bottom patterns (learn specific cue-criterion relationships), then how water temperatures influence sonar image (another cue-criterion relationship), and only after that, how ocean bottom pattern and water temperatures interact and how that influences the sonar image (see e.g. [www.mosaichydro.com](http://www.mosaichydro.com)). Another example involves training of Air Traffic Controllers: first they learn how to separate aircraft on the basis of heading only (i.e. not dealing with altitude, or speed), then they practice separating

aircraft on the basis of altitude and incorporate that with heading in their practice, after that the learn about airspeed, groundspeed, and how that relates to separating aircraft and lastly they practice separating aircraft on the basis of all aforementioned attributes (Oprins, 2008). The data presented here, however, show that despite the beneficial effects of proactive prompts in blocked schedules, such training programs would be well-advised to present the trainee with a random sequence of cases, leaving it to the trainee to identify and categorize cues and cue-criterion relationships.

## References

- Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M., Tamim, R., & Zhang, D. A. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage one meta-analysis. *Review of Educational Research*, *78*, 1102 - 1134.
- Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*, 147–179.
- Alexander, P. A., Schallert, D. L., Reynolds, R. E., (2009). What Is Learning Anyway? A Topographical Perspective Considered. *Educational Psychologist*, *44*, 176-192.
- Angeli, C. & Valanides, N., (2008). Instructional effects on critical thinking: Performance on ill-defined issues. *Learning and Instruction*, *19*, 322-334.
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (in press). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology, Research and Development*. Retrieved January 20, 2008, from <http://dx.doi.org/10.1007/s11423-007-9074-1>.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.
- Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L-G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological Perspectives* (pp. 109-133). New York: Psychology Press.
- Björkman, M. (1965). Studies in predictive behavior: Explorations into predictive judgments based on functional learning and defined by estimation, categorization, and choice. *Scandinavian Journal of Psychology*, *6*, 129-156.

- Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. *Quest, 50*, 266-293.
- Bransford, J. D. & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad and P. D. Pearson (Eds.), *Review of Research in Education*, 24, 61-100. Washington, D.C.: American Educational Research Association.
- Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75-114). Amsterdam, The Netherlands: North-Holland Elsevier.
- Brehmer, B. (1972). Cue utilization and cue consistency in multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 8*, 286-296.
- Brehmer, B., & Joyce, C. R. B. (Eds.) (1988). *Human judgment: The SJT view*. Amsterdam, The Netherlands: North-Holland Elsevier.
- Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist, 52*, 399-413.
- Brunswik, E., & Herma, H. (1951). Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology, 41*, 281-290.
- Carlson, R. A. (1989). Processing nonlinguistic negation. *American Journal of Psychology, 102*, 211-224.
- Carlson, R. A., & Schneider, W. (1989). Acquisition context and the use of causal rules. *Memory and Cognition, 17*, 240-248.
- Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 517-526.

- Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 484-496.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Erlbaum.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Selfexplanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.
- Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.
- Clark, A., Anderson, R. C., Kuo, L. J., Kim, J., Archodido, A., & Nguyen-Jahiel, K. (2003). Collaborative reasoning: Expanding ways for children to talk and think in school. *Educational Psychology Review*, *15*, 181-198.
- Cohen, M. S. (1986). An expert system framework for non-monotonic reasoning about probabilistic assumptions. In J. F. Lemmer & L. N. Kanal (Eds.), *Uncertainty in artificial intelligence* (pp. 279-293). Amsterdam: North Holland
- Cohen, M. S. & Freeman, J. T. (1998). Understanding and enhancing critical thinking in recognition-based decision making. In R. Flin & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications*. Avebury Aviation.
- Cohen, M. S., Freeman, J. T., & Thompson, B. B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. In J.A. Cannon-Bowers & E. Salas

- (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 155-190). Washington, DC: American Psychological Association.
- Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Meta-recognition in time stressed decision making: Recognizing, critiquing, and correcting, *Human Factors*, 38, 206-219.
- Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning questions effect: The role of dialogue and deep-level reasoning questions during vicarious learning. *Cognition and Instruction*, 24, 565-591.
- Cross, E. S., Schmitt, P. J., & Grafton, S. T. (2007). Neural substrates of contextual interference during motor learning support a model of active preparation. *Journal of Cognitive Neuroscience*, 19, 1854 - 1871.
- Davis, E. A. (2003). Untangling dimensions of students' beliefs about scientific knowledge and science learning. *International Journal of Science Education*, 25(4), 439-468.
- Davis, E. A., & Linn, M. C.. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, 22, 819-837.
- De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G., (2005). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology*, 19, 167-181.
- De Croock, M. B. M., Van Merriënboer, J. J. G., & Paas, F. (1998). High vs low contextual interference in simulation based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior*, 14, 249-267.
- Detterman, D. K., & Sternberg, R. J. (Eds.) (1993). *Transfer on trial: intelligence, cognition, and instruction*. Norwood, NJ: Ablex Publishing.

- Dhmi, M. K., Hertwig, R., & Hoffrage, U. (2004). The Role of Representative Design in an Ecological Approach to Cognition. *Psychological Bulletin*, *130*, 959-988
- Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning*. Utrecht, The Netherlands: Lemma BV.
- Douglas, M., Hosokawa, M., & Lawler, F. (1988). *A practical guide to clinical teaching in medicine*. New York: Springer.
- Drillings, M., & Serfaty, D. (1997). Naturalistic decision making in command and control. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 71-80). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.
- Ennis, R. (1989). Critical thinking and subject-specificity: clarification and needed research. *Educational Researcher* *18*, 4-10.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA, The California Academic Press: 19.
- Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics text: Differences in study processes between good and poor performers. *Cognition and Instruction*, *7*, 41-54.
- Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision-making: A test of instruction based on the recognition / metacognition model. *Proceedings of the 3rd International Command and Control Research and Technology Symposium 1996*. Retrieved May 12, 2005 from [www.cog-tech.com/Publications/TrainingPubs.htm](http://www.cog-tech.com/Publications/TrainingPubs.htm).

- Frijters, S., Ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instruction, 18*, 66-82.
- Gagné, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39*, 377-385.
- Guadagnoli, M. A., & Lee, T. D. (2004) Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*, 212-224.
- Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking* (4th ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71*, 438-456.
- Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). *Human judgment and decision making: Theories, decisions, and procedures*. New York: Praeger Publications.
- Helsdingen, A.S., Van den Bosch, K., Van Gog, T. & Van Merriënboer, J.J.G. (2010). The effects of critical thinking instruction on learning complex judgment and decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, Advance online publication. doi:10.1177/0018720810377069*
- Helsdingen, A.S., Van Gog, T., & Van Merriënboer, J.J.G. (2010). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction, Advance online publication. doi:10.1016/j.learninstruc.2009.12.001*
- Hoffman, P. J., Earle, T. C., & Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptions of feedback. *Organizational Behavior and Human Performance, 27*, 75-102.

- Hogarth, R. M. (1980). *Judgment and choice*. New York: John Wiley and Sons Ltd.
- Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review*, *71*, 42-60.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649-667.
- Jamieson, B. A., & Rogers, W. A. (2000). Age-related effects of blocked and random practice schedules on learning a new technology. *The Journals of Gerontology series B: Psychological Sciences and Social Sciences*, *55*, 343-353.
- Jelsma, O., Van Merriënboer, J. J., & Bijlstra, J. P. (1988). The ADAPT design model: Towards instructional control of transfer. *Instructional Science* *19*(2), 89-120. a
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude treatment approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657-690.
- Knutzon, J. S., Walter, B. E., Sannier, A. V., & Oliver, J. H. (2004). An immersive approach to command and control. *Journal Of Battlefield Technology*, *7*, 37-42.
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219-224.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, *28*, 16-25,
- Lee, A. Y., & Hutchison, L. (1998). Improving learning from examples through reflection. *Journal of Experimental Psychology: Applied*, *4*, 187-210.
- Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 730-746.

- Lee, T. D., & Simon, D. A. (2004). Contextual interference. In A. M. Williams & N. J. Hodges (Eds.), *Skill acquisition in sport: Research, theory and practice* (pp. 29-44). London, UK: Routledge.
- Lin, C. J., Fisher, B. E., Winstein, C. J., Wu, A. D., Gordon, J. (2008). Contextual interference effect: Elaborative processing or forgetting–reconstruction? A post hoc analysis of transcranial magnetic stimulation–induced effects on motor learning. *Journal of Motor Behavior*, *40*, 578–586.
- Magill, R. A., & Hall, K. G. (1990). A review of contextual interference effect in motor skill acquisition. *Human Movement Science*, *9*, 241-289.
- Morgan, N., & Saxton, J. (1991). *Teaching, questioning and learning*. London, UK: Routledge.
- Naidu, S., & Bernard, R. M. (1992). Enhancing academic performance in distance education with concept mapping and inserted questions. *Distance Education*, *13*, 218-233.
- Nickerson, R. S. (1987). Why teach thinking?, in J. B. Baron and R. J. Sternberg (Eds.), *Teaching Thinking Skills: Theory and Practice*. (pp. 27-37). New York: W. H. Freeman.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- Oprins, E. (2008). *Design Of A Competence based assessment system for air traffic control training*. Doctoral thesis. University press Maastricht.
- Orlich, D. C., Harder, R. J., Callahan, R. C., Kauchak, D. P., & Gibson, H. W. (1994). *Teaching strategies: A guide to better instruction* (4<sup>th</sup> Ed.). Toronto, Canada: Heath and Company.
- 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., & 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.
- Perkins D. N. & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher 18*, 6–95.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1-29.
- Richland, L.E., Bjork, R.A., Finley, J.R., & Linn, M.C. (2005). Linking cognitive science to education: generation and interleaving effects. In B. G. Bara, L. Barsalou and M. Bucciarelli (Eds.) *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation psychology* (pp. 182-193). Ames, IA: Iowa State University Press.
- Rowntree, D. (1992). *Exploring open and distance learning*. London, UK: Kogan Page.
- Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497-510.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.
- Schneider, V. I., Healy, A. F., & Bourne, L. E. Jr. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 78-90). Mahwah, NJ: Erlbaum.

- Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention and transfer. *Journal of Memory and Language*, 46, 419-440.
- Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (1995). The effects of contextual interference on the acquisition and retention of logical rules. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 95-131). Thousand Oaks, CA: Sage Publications.
- Schnotz, W. & Kürschner, C. (2007). A reconsideration of cognitive load theory, *Educational Psychology Review*, 19, 469–508.
- Schworm, S. & Renkl, A. (2007) Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99, 285-296.
- Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention and transfer of a motor skill. *Journal of experimental psychology: Human learning and memory*, 5, 179-187.
- Siegel, H. (1991). The generalizability of critical thinking. *Educational Philosophy and Theory*, 23, 18-30.
- Simon, D. A. (2007). Contextual interference effects with two tasks. *Perceptual and Motor Skills*, 105, 177-183.
- Singley, M. K. & Anderson, J. R. (1989). *Transfer of Cognitive Skill*. Cambridge, MA: Harvard University Press.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, 57, 193-216.
- Smedslund, J. (1955). *Multiple-Probability Learning: An Inquiry Into The Origins Of Perception*. Oslo, Norway: Oslo University Press.

- Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*, 701-726.
- Sweller, J., (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296.
- Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education*, *40*, 185-200.
- Van den Boom, G., Paas, F., Van Merriënboer, J.J.G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, *17*, 532-548.
- Van den Boom, G., Paas, F., Van Merriënboer, J.J.G., & Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*, 551-567.
- Van den Bosch, K., & Helsdingen, A. S., (2006). Critical thinking in tactical decision games-training. In: M. J. Cook, J. M. Noyes, & Y. Masakoski (Eds.), *Decision making in complex systems* (pp. 213-222). Aldershot, England: Ashgate
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*, 154-164.
- Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, *11*, 237-244.

- Van Merriënboer, J. J. G., & Paas, F. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior*, 6, 273-289.
- 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.
- Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30, 173-187.
- Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychological Bulletin & Review*, 9, 185-211.
- Zechmeister, E. & Johnson, J. (1992). *Critical Thinking: A Functional Approach*. Pacific Grove, CA: Brooks/Cole.

Appendix A: Instruction to participants, translated (original in Dutch)

If we want to prepare professionals for the complexities of operational decision making, than training should aim to (a) expand and elaborate relevant knowledge of the domain, and (b) practice in situation assessment and judgment in complex situations. In order to acquire domain knowledge, it is important that a decision maker is presented with adequate practice scenarios.

However, to effectively expand and deepen the domain knowledge, it is not enough to just practice those scenarios. The learning is enhanced when the various cases are being studied from different perspectives and when the consequences of various options are thoroughly evaluated. The accumulation of domain knowledge requires a long period of intensive, focussed and reflective practice. That principle holds true for most if not all professions. One doesn't become a grandmaster by merely playing many games of chess – it requires focussed study of positions and other people's matches. Neither can one become a renowned piano player by just playing many sonatas; it requires intensive practise of piano techniques and study of musical invariants. The same principle applies to many professional decision-making, e.g. military command and control: one doesn't become an expert by merely running a lot of scenarios – what one learns depends on the profundity of the practice exercises. We have developed a training method that is based on the way expert decision-makers work in complex situations and that stimulates an active, focussed and reflective study of training scenarios. That method is called "*Critical Thinking*".

In the *Critical Thinking Training* the emphasis is on the application of expert strategies. What are the strategies of an expert and in what way do they differ from those of non-experts? If less experienced decision makers are confronted with an unknown or complex situation, they often base their assessment of the situation on only a few cues. Often they do not recognize how

the various cues are or could be related as they are therefore often inclined to come to conclusions through isolated cues. When experienced decision makers are confronted with an unknown or complex situation, they use their domain knowledge to get to a first impression of the problem. As a next step they continue to actively search for additional information to help them complement, refine or reassess their initial view. They also try to verify whether their assumptions are and look for coherency in the available data.

In short, experts are often capable to act at two different levels at the same time. Of course they act at the task execution level, but on top they also manage to observe their own approach to the task, to guard it and to guide it. That allows them to look beyond the superficial characteristics of the situation they are in. Through reflection on the situation and their own way of handling it, they achieve a better understanding of the situation and risks involved. Their actions therefore, often yield better results. The approach by experts is the model for our *Critical Thinking Training*.

Our training is focussed on training trainees to acquire an approach that experts have learned through experience. An obvious challenge we had to overcome is that experts do not have a separate work instruction on how to become experts, but that getting experience is an integral part of 'doing the job'. To transfer the expert approach to non-experts, we have designed the *Critical Thinking Methodology*. This method consists of four distinctive steps:

*Creating a story*: A story is a comprehensive assessment of the situation, in which all existing evidence is incorporated and explained and assumptions are made about uncertain aspects of the situation. Past, present and future are addressed in the story. The purpose of story building is to keep participants from assessing situations solely on isolated events. Instead, participants are taught how they can integrate the available information into its context.

*Testing a story:* Participants are instructed how to identify inconsistency and uncertainty, and how to adjust or refine their story by deliberate testing. They have to correct these problems by collecting more data, retrieving knowledge from memory, making assumptions about the missing piece of the story, or by resolving conflicts in the argumentation.

*Evaluating a story:* After a story is constructed and tested for gaps, it should be evaluated for its plausibility. The decision maker has to take a step back, identify critical or hidden assumptions and play the devil's advocate by falsifying these assumptions, that is, explaining how an assumption can be false and build an alternative story.

*Quick test:* Critical thinking is not always appropriate. Decision makers have to evaluate the time available and the consequences of their actions. In stressful situations such as those often encountered by professional decision makers, there is usually little time to spare. The decision makers should act immediately *unless* the risk of delay is acceptable, the costs of errors are high (i.e., when lives are at stake), and/or the situation is non-routine or problematic (i.e., complex or novel problems).

Appendix b: Case description transfer task (translated from Dutch)<sup>2</sup>

Seventeen-year old Kai Overduin takes his father's red BMW to drive to his friend Gerard's house. Kai and Gerard then drive to town, park the car close to a bar and sit outside on the sidewalk terrace. Kai starts of drinking sodas, like he always does when he 'borrows' his father's car, but after three Pepsi's, he orders a beer. Just one. And after a while another one, and just before they go just one more. Around six o'clock the two boys drive back to Kai's place because his father can come home any minute now, and the car has to back at the house before that. Kai feels a little weird driving after three beers, because he is not used to drink and drive, so he's careful not to drive too fast.

---

<sup>2</sup> Other case descriptions can be found at <http://files.me.com/anne.h/a0thyn>

**Table 1***Schematic Overview of the Cues, Cue Values and Subsequent Priorities*

<b>Training and post-testcases (crimes)</b>			
<b>Condition of victim</b>	<b>Weapon</b>	<b>Nature of crime</b>	<b>Priority</b>
Dead	Firearm	Burglary	62
		Violence	91
	No firearm	Burglary	58
		Violence	72
Injured	Firearm	Burglary	29
		Violence	43
	No firearm	Burglary	10
		Violence	39
<b>Transfer cases (traffic offenses)</b>			
<b>Nature of the offense</b>	<b>Condition of Offender</b>	<b>History</b>	<b>Priority</b>
Speeding	Drunk	First offence	84
		Recidivist	94
	Not drunk	First offence	64
		Recidivist	74
Driving without insurance	Drunk	First offence	32
		Recidivist	40
	Not drunk	First offence	12
		Recidivist	20

Table 2

*Critical Thinking Prompts*

<b>Proactive prompts</b>	<b>Reflective prompts</b>
Are there any similarities between the following two cases? What are they? And what is different between these cases?	Were there any similarities between the last two cases? What were they? And what were the differences?
The following two cases will differ from the former two cases on one specific aspect. Can you tell me what that is? And does it have an effect on the priority of a case?	The last two cases were different from the two preceding cases on one specific aspect. Can you tell me what that was? And did it have an effect on the priority of a case?
Do you consider some aspects to be more important than others when estimating a case's priority?	Do you consider some aspects to be more important than others when estimating a case's priority?