# Cultivating critical thinking

*The effects of instructions on economics students' reasoning*


Anita Heijltjes

# Cultivating Critical Thinking:

## The Effects of Instructions on Economics Students' Reasoning

ISBN: 978-90-8891-774-5



Cover design: Martijn van Dalen

Printed & Lay Out by: Proefschriftenmaken.nl, Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, ’s-Hertogenbosch

# Cultivating Critical Thinking:
## The Effects of Instructions on Economics Students' Reasoning

# Cultiveren van kritisch denken:
## De effecten van instructies op het redeneren van economie studenten

**Proefschrift**

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H. A. P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 30 januari 2014 om 15.30 uur

door

**Antonia Elisabeth Gerarda Heijltjes**
geboren te Zevenaar

ERASMUS UNIVERSITEIT ROTTERDAM

Contents

**Promotiecommissie**

| | |
|---|---|
| Promotoren | Prof.dr. F. Paas<br>Prof.dr. T. van Gog |
| Overige leden | Prof.dr. J. J. Beijshuizen<br>Prof.dr. R. M. J. P. Rikers<br>Prof.dr. J. J. G. Van Merriënboer |

# Contents

Contents

# 1 General Introduction

Contents

John Dewey stated in 1910 that 'habits of reflective thoughts' are 'not a gift of nature' but can and should be 'cultivated' by education (Dewey, 1997). His concept of 'habits of reflective thoughts' is closely related to more contemporary concepts of *critical thinking*, which has been comprehensively defined as "purposeful, self-regulatory judgment, which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." (Facione, 1990, p. 2). Despite the ideal –advocated by scholars, administrators, and educators- that education should foster students' critical thinking (Bok, 2006; Facione, 1990; Halpern, 1999; Siegel, 1989), critical thinking skills of many undergraduates do not seem to improve during their college years (Arum & Roksa, 2011), and subsequently graduate with poor thinking skills (Flores, Matkin, Burbach, Quinn, & Harding, 2012). This is quite problematic, because the development of critical thinking skills has been inextricably linked to being successful in the complex 21$^{st}$ century world (Pellegrino & Hilton, 2012). Critical thinking enables students to make sound, logical, and unbiased decisions (Facione, 1990; Paul, 1990), facilitates lifelong learning (Halpern, 1998), cultivates responsibility and competencies for good democratic citizenship (Nussbaum, 2006), and leads to better learning and transfer (e.g., Helsdingen, Van Gog, & Van Merriënboer, 2011). So why, despite the many grounds to warrant active efforts of educators to foster students' critical thinking skills, do many higher education graduates show unsatisfactory improvement in critical thinking? One reason is that it is rarely explicitly taught (Davies, 2013; Paul, 2005), as has been noted in business and economics education (Jones, 2007; Smith, 2003), even though research has shown that critical thinking skills will not develop as a by-product of education (e.g., Abrami et al., 2008; Arum & Roksa, 2011; Bangert-Drowns & Bankert, 1990; Marin & Halpern, 2011). In educational research many attempts have been made to develop and test critical thinking teaching strategies, and although a meta-analysis showed that teaching general critical thinking strategies as an independent track within a content course leads to the best outcomes (Abrami et al., 2008), it is still unclear how critical thinking skills can be best taught (e.g., Niu, Behar-Horenstein, & Garvan, 2013). Moreover, one essential aspect of critical thinking, namely the ability to enagage in unbiased reasoning, has received little attention in educational research thus far (West, Toplak, & Stanovich., 2008).

The aim of the studies presented in this dissertation was to test the impact of different critical thinking instructions on this essential aspect of critical thinking, that is, the ability to enagage in unbiased reasoning. The main aim was to investigate

Contents

the effects of general explicit critical thinking instructions (as compared to implicit instruction or no instruction) and the effects of practice in a domain context (with or without prompts to foster deeper reasoning), on the acquisition of reasoning skills. As dispositions are an essential part of the critical thinking concept (Facione, 1990), the second aim was to explore the role of students' thinking dispositions in reasoning and whether they interact with the effects of instructions. The third aim was to explore the role of confidence and mental effort in reasoning prior to and after instruction.

Before providing an overview of the studies this dissertation contains, the theoretical background of critical thinking instruction will be discussed.

## Critical Thinking Defined

The concept of 'reflective thought' refers to an "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusion to which it tends" (Dewey, 1997, p. 6). This concept has evolved over the past century into many definitions of 'critical thinking'. An expert Delphi Panel of the 'American Philosophical Association' (APA; Facione, 1990) characterized critical thinking –mentioned above- as "purposeful, self-regulatory judgment, which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." (Facione, 1990, p. 2). The concept of critical thinking has been elaborated in great detail by the APA Delphi Panel, and includes dispositions and cognitive skills.

Dispositions refer to 'habits of mind' such as "open-mindedness concerning divergent worldviews, flexibility in considering alternatives and opinions, and willingness to reconsider and revise views when change is warranted" (Facione, 1990, p.25). At the highest level, cognitive skills refer –amongst others- to skills such as interpretation, evaluation, and inference (Facione, 1990).

Besides the APA concept of criticical thinking, many other definitions have been proposed (Ennis, 1985; Halpern, 1998; Kuhn, 1999; Paul, 1990; Siegel, 1980). Most of the definitions show much overlap, although there are differences in whether critical thinking skills are viewed as domain general or domain specific. For example Ennis (1985) and Paul (1985) argued that critical thinking is domain general, that is, independent of specific disciplines, whereas McPeck (1990) argued that criteria for applying and assessing critical thinking have to be derived from specific topics, subjects, fields or domains (i.e., domain specific). This distinction, about which there is ongoing debate (e.g., Davies, 2013; Moore, 2011), also implies different views on how to instruct critical thinking. For instance, according to Ennis (1989) general critical thinking

Contents

principles can be taught that are then applied to a specific domain, whereas McPeck (1990) argues that critical thinking cannot be instructed independently from a specific domain.

Taking a domain-general view, a number of researchers have linked critical thinking more strongly to the concept of rationality and the ability to engage in unbiased reasoning (Siegel, 1989; Stanovich, 2011; West et al., 2008), which is also the focus of this dissertation. According to Siegel (1989) "to be a rational person is to believe and act on the basis of reasons" (p. 21), and education should aim to foster rationality, by focusing on the reasons on which assessments, judgments, and actions are based. By identifying thinking errors such as reasoning biases, an individual's rationality can be assessed (Gilovich & Griffin, 2002; Stanovich, 2011; Tversky & Kahneman, 1981).

## Biased Reasoning from a Dual Processing Perspective

The literature on critical thinking in educational research has focused on a wide variety of topics in the past decades (Angeli & Valanides, 2009), but has paid little attention to the aspect of biased reasoning (West et al., 2008). Moreover in the 'heuristics and biases' literature thinking fallacies have been comprehensively studied, but effective debiasing techniques and strategies have received much less attention (Klaczynski, 2006; Klaczynski, Gordon, & Fauth, 1997; Larrick, 2004). Studying biased reasoning and how to prevent it is an important area of research (Larrick, 2004) that complements traditional research on critical thinking (West et al., 2008). Biases seem inherent to human reasoning and decision making.

For instance, according to Kahneman and Tversky (1973) people are often overly confident, fear losses more than they value gains, are influenced by frames and prior beliefs when making judgments, and favor personal or contextual cues over more representative statistical evidence (e.g., neglect relevant base-rates) when making judgments. Moreover, people tend to accept a conclusion following from premises as valid if it is believable, regardless of its logical validity (e.g., Evans, Handley, & Harper, 2001), and tend to select information that matches the wording of a propositional statement about which they are reasoning while neglecting logically relevant information (Evans, 2003).

The ability to engage in unbiased reasoning is crucial for decision-making in complex and high-risk professions such as economics (Smith, 2003). In rapidly changing and complex business environments, biased reasoning can result in erroneous decisions with severe consequences. For instance, Ireland's banking crisis in 2008 has recently been associated with a number of cognitive biases leading to the underestimation of risks by stake-holders (Lunn, 2011). Research among financial market professionals showed that they forecasted earning per share too optimistically (De Bondt & Thaler, 2002), and stockbrokers have been demonstrated to engage in irrational thinking, drawing invalid conclusions guided by prior knowledge and beliefs instead of logical reasoning (Knauff, Budeck, Wolf, & Hamburger, 2010).

Biases and fallacies can be explained by dual process models, which state that two types of cognitive processes are involved in reasoning (Evans, 2003, 2008, 2011; Stanovich, 2011). Type 1 processes have an, automatic nature, involve little reflection, and impose a relatively low load on working memory. Decision-making using Type 1 processes, is based on past experiences, which is useful and efficient in many routine situations. However, because of its automaticity, it might result in biased thinking in other, non-routine situations. Type 2 processes are deliberate, and sequential in nature, and impose a higher load on working memory than Type 1 reasoning (Evans, 2012). Type 2 reasoning can overrule these automatic responses (Type 1 pocesses) by explicit reasoning efforts.

According to Stanovich (2011), Type 2 reasoning can be subdivided into algorithmic processes and reflective processes. Reflective processes are associated with beliefs, cognitive styles, goals, and epistemic values and thinking dispositions (see section on Dispositions). Algorithmic processes are associated with analytic and inhibitory operations and with decoupling beliefs from evidence (Stanovich, 2011). Thus, the 'algorithmic mind' has the ability to override Type 1 processes by applying knowledge of inferential rules and strategies of rational thought (e.g., probabilistic reasoning, causal reasoning and logic). Biased reasoning arises when Type 2 reasoning fails to override Type 1 reasoning in situations in which Type 1 reasoning is insufficient (Stanovich, 2011). Such failures to override Type 1 reasoning might be due to a lack of declarative knowledge, or, when such knowledge is present, due to insufficiently developed strategies such as hypothetical thinking (i.e., engaging in cognitive simulations to consider alternative responses or possibilities by decoupling secondary representations from primary representations; Stanovich, 2011).

The question addressed in this dissertaton is whether a lack of declarative knowledge and knowledge of reasoning strategies might be counteracted by explicit instructions, thereby reducing biases in reasoning.

## Instructions to Avoid Biased Reasoning

As mentioned in the beginning of this Introduction, the educational literature on teaching critical thinking (i.e., not specifically on reasoning), has shown that explicit, general critical thinking instruction combined with the opportunity to integrate the general principles that were taught with domain-specific subject matter, works best

Contents

(e.g., Abrami et al., 2008; Bangert-Drowns & Bankert, 1990). For example, Marin and Halpern (2011) compared the effectiveness of critical thinking instruction woven into the content matter with additional explicit instruction by means of supplementary critical thinking lessons. Participants who received explicit instruction outperformed those who did not, and moreover, showed transfer of critical thinking skills to everyday situations.

The question is, however, whether this holds also for instructions intended to 'debias biases' as it was put by Larrick (2004). Would rational thinking in terms of overriding Type 1 reasoning be trainable, and if so, what type of instruction would be most beneficial? The available research reveals some evidence for positive effects of explicit training (Larrick, Morgan, & Nisbett, 1990; Macpherson & Stanovich, 2007) but raises the question of whether that should be combined with opportunities for practice. Explicit critical thinking instructions aimed at unbiased reasoning have been shown to be beneficial in prior research, but the findings are mixed; bias reasoning has been found to improve from explicit instruction (Fong, Krantz, & Nisbett, 1986; Macpherson & Stanovich, 2007) or example training (Fong et al.,1986), or might only improve from a combination of instructions and example training (Cheng, Holyoak, Nisbett, & Oliver, 1986).

Although findings regarding teaching reasoning are mixed, combined with the findings on teaching other aspects of critical thinking (Abrami et al., 2008), it seems that only providing general instructions might be insufficient, and that this might have to be combined with practice in a domain-specific context. That would allow students to integrate general rules, to combine and cluster information in a meaningful way, and to mentally reorganize their knowledge in such a way that it can be accessed when needed (Bransford, Sherwood, Vye, & Rieser, 1986). Perkins and Salomon (1988) emphasized that bringing together general and specific knowledge maximizes the likelihood of obtaining transfer of critical thinking skills and reasoning.

Assuming that both explicit, general instruction and opportunities for applying the reasoning principles that were instructed in a domain-specific context are required for the acquisition of reasoning skills, the question is how to make the most of practice. If practice allows for better integration and elaboration of the information from the instructions, then would prompting students' thinking during practice support those processes and have added value for acquiring reasoning skills? Two different prompting strategies were studied in this dissertation: self-explanation prompts and activation prompts.

According to Roy and Chi (2005), self-explanation is a domain-general constructive activity of explaining instructional materials to oneself, which engages students in active and meaningful learning while they effectively monitor their understanding.

Self-explaining presumably helps students in identifying comprehension failures, integrating new information with prior knowledge and repairing faulty knowledge (Roy & Chi, 2005). According to Lombrozo (2006), explaining is an especially useful approach for reasoning. She argues that generating an explanation of why a claim might be true enables one to assess the probability of that claim in light of prior beliefs and subsequently encourages students to re-describe materials, revise beliefs, and identify relevant principles. The self-explanation process can be constrained by prior knowledge and beliefs and therefore explanations are in particular effective if novel information is related to prior beliefs. Prompting self-explanation has been found to foster proper use of available knowledge (Roy & Chi, 2005), problem solving and transfer (e.g., Aleven & Koedinger, 2002; Renkl, 2005; Rittle-Johnson, 2006) and revising of beliefs (Williams & Lombrozo, 2010). Although various studies have shown the benefits of prompting self-explanation as an instructional strategy (e.g., Hilbert, Schworm, & Renkl, 2004; Rittle Johnson, 2006; Schworm & Renkl, 2006, 2007), there are also studies that found little or no beneficial effects of prompting self-explanation on learning (Matthews & Rittle-Johnson, 2009; McEldoon, Durkin, & Rittle-Johnson, 2012). Looking specifically at effects of self-explaining on reasoning tasks that are prone to bias (i.e., base-rate, conjunction, and gambler fallacy tasks), Austin, Gregory, and Chiu (2008) showed that pharmacy students who were prompted to provide a brief written explanation and justification of how they arrived at their answer, demonstrated a significantly better performance than students who did not receive these prompts. Because an essential aspect of unbiased reasoning (Type 2 processes) is to decouple prior beliefs from evidence, prompting self-explanation during practice of reasoning tasks might have an additional impact on the acquisition of reasoning skills.

Activation prompts function in a different manner; they are assumed to draw attention away from salient but non-relevant information and to activate relevant but less salient information instead. Moutier, Angeard, and Houdé (2002), as well as Moutier and Houdé (2003), showed that inhibiting a dominant response can lead to activation of a more relevant response, thereby improving reasoning performance. For example on conjunction tasks, participants were taught to inhibit the misleading scheme (the conjunction) in order to activate the relevant scheme (the probability). The underlying assumption is that reasoning biases are caused by an inhibition failure in working memory and not due to a lack of understanding of principles or a lack of competence. Stanovich and Stanovich (2010), however, stated that suppressing the initial response (Type 1 reasoning) is only helpful when a better response is available to substitute for it. Therefore, the use of prompts to activate the 'weaker' cues instead of prompts to suppress dominant cues might be more effective (i.e., evoke an attention shift from irrelevant to relevant information; Miyake, Friedman, Emerson, Witzki, & Howerter,

Contents

2000). Because such 'activation prompts' during practice would encourage an attention shift to relevant task aspects, it was expected that they might have an additional impact on the acquisition of reasoning skills.

The potential benefits of explicit, general instructions, either combined with practicing reasoning tasks on a domain-specific business case or not, as well as the effectiveness of prompting self-explanation, was investigated in the studies presented in Chapters 2, 3 and 4. The effectiveness of activation prompts was explored in the studies reported in Chapter 2 and 3. In addition the study presented in Chapter 2 compared the effects of implicit critical thinking instruction in a course on argumentation with explicit critical thinking instruction. As practicing tasks often requires additional time on task compared to other methods, the study reported in Chapter 4 compared instruction plus practice with an instruction only but combined with a deep processing strategy that used the same amount of time.

## Dispositions

Thinking dispositions have been identified as essential component of the concept of critical thinking (Facione, 1990) and there is empirical evidence that scores on dispositional constructs are related to biased reasoning (e.g., West et al., 2008). For instance dispositions such as 'actively open-minded thinking' (i.e., AOT: the active search for evidence against one's own beliefs, plans, or goals and the ability to weigh available evidence fairly; Baron, 2008) and 'need for cognition' (NFC, i.e., "the tendency for an individual to engage in and enjoy thinking"; Cacioppo, Petty, & Kao, 1984, p. 306) have been found to predict performance on several reasoning tasks. For instance, individuals with higher AOT scores performed better on argument evaluation (Stanovich & West, 1997) and avoided unsophisticated reiteration in argument (Sá, Kelley, Ho, & Stanovich, 2005). West et al., (2008) found that scores on AOT and NFC were unique predictors on a composite score of heuristics and biases tasks (e.g., causal and non-causal base-rate tasks, Wason selection tasks, and Framing tasks) and belief bias syllogisms. NFC is characterized by cognitive motivation, and individuals with higher NFC scores were able to recall more of the information to which they were exposed, were influenced more by the quality of arguments, generated more task-relevant thoughts, and put more effort into cognitive tasks than individuals with lower NFC scores (for a review, see Cacioppo, Petty, Feinstein, & Jarvis, 1996). In all empirical chapters of this dissertation, it is investigated whether the association between dispositions (i.e., AOT and NFC) and reasoning performance can be replicated, and whether dispositions have an impact on the effects of critical thinking instructions (e.g., would students with higher scores on dispositions learn more from the instructions than students with lower scores on dispositions?).

## Confidence and Cognitive Effort in Reasoning

Both confidence in responses on reasoning tasks and mental effort invested in those tasks seem to be associated with the extent to which people engage in Type 1 and Type 2 reasoning (Evans, 2012; Thompson, 2009). As for mental effort, a crucial difference between Type 1 and Type 2 thinking processes is the extent to which working memory is involved: Type 1 is automatic, imposing little if any load on working memory, while Type 2 is deliberate and imposes high working memory load (Evans, 2008; 2012).

As for confidence, it has been found that the fluency with which an answer comes to mind can lead to a 'feeling of rightness', which determines to what extent Type 2 reasoning is subsequently engaged in (Thompson, 2009). For example, participants who had a weak feeling of rightness about an initial judgment spent more time on reaching an answer, and were more likely to change their answer than participants with a strong feeling of rightness (Thompson, Prowse Turner, & Pennycook, 2011). In the context of social judgment, research has shown that individuals exert only as much mental effort as necessary to reach a satisfying level of confidence in their judgments, which has been called the sufficiency principle (Chen & Chaiken, 1999). According to Chen and Chaiken, people are 'economy minded' and balance between minimizing their effort investment and feeling sufficiently confident. If a desired level of confidence has been reached, people tend to put no more effort into their judgment. This is problematic, as it has often been found that confidence in reasoning tasks and accuracy of performance on those tasks are poorly related (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Griffin & Tversky, 2002). Often, good and poor reasoners show equal confidence levels, despite large differences in accuracy (Shynkaruk & Thompson, 2006). Presumably, this is due to the aforementioned mechanisms (i.e., fluency and sufficiency). Therefore on an initial assessment of reasoning skills, a poor relation between performance and confidence can be expected, and confidence and effort are likely to be negatively related.

A more interesting question, however, is whether and how instruction would affect the relation between performance, invested mental effort and confidence in reasoning tasks.

Regarding performance and confidence, on the one hand, instructions might lead to higher confidence as well as higher accuracy, but on the other hand, in the case of initial overconfidence, the level of confidence might remain the same or become lower while accuracy improves. For instance, Koriat, Lichtenstein, and Fischhoff (1980) found that participants, who were instructed to search for and attend to contradictory evidence,

Contents

lowered their confidence and improved their judgments. Prowse Turner and Thompson (2009) on the other hand, found that training affected participants' accuracy but had little or no effect on their confidence levels.

Regarding performance and mental effort, educational research has shown that knowledge of strategies gained through instructions might reduce the cognitive load imposed by the task and therefore students have to invest less mental effort to achieve higher performance on a subsequent test (Paas, Renkl, & Sweller, 2003). Whether this would also be the case on reasoning tasks is unclear, because the instructions in this case are intended to evoke a shift from Type 1 to Type 2 processing, which is more effortful than Type 1 processing (Evans, 2011), and might therefore lead to more effort investment on a subsequent test. It is also important here to distinguish between tasks that are instructed (and practiced) and tasks that are not; possibly, higher performance, lower effort, and equal/higher confidence will be seen in tasks that have been instructed (given that the strategy for handling such tasks is known from the instructions), whereas equal/higher performance, higher effort, and equal/lower confidence might be seen in tasks that were not instructed (as having received instructions on other tasks might evoke Type 2 thinking on those tasks, but it is not yet known how to handle those tasks).

The questions addressed in this dissertation are how mental effort is related to confidence and reasoning performance, and whether and how instructions (with practice) affect confidence (Chapter 4) and invested mental effort (Chapter 2 and Chapter 4).

## Overview of the Dissertation

This dissertation consists of three empirical studies on the impact of different kinds of instructional conditions and individual differences in dispositions on critical thinking in terms of the ability to engage in unbiased reasoning. All studies were conducted at a Dutch University of Applied Sciences. Chapter 2 presents an experiment in which second year students of an Economics and Business Education department ($N$ = 141) participated, investigating the impact of different kinds of critical thinking instructions and thinking dispositions on their reasoning performance. Reasoning performance was operationalized as performance on reasoning tasks prone to bias. Regarding instruction, it was investigated whether a) critical thinking would improve through instructions provided implicitly in a regular course context or whether explicit critical thinking instruction was needed, b) if so, whether explicit, general instruction needed to be combined with practice in a domain-specific context, and c) whether self-explanation and activation prompts during practice would have added value. Moreover it was investigated whether participants with higher scores on dispositions (i.e., Actively

Open-minded Thinking AOT and Need for Cognition) would score better on the pre-test as had been found in prior research (West et al., 2008), and, more interestingly, whether dispositions would interact with the effect of instructions (e.g., whether participants with higher scores on dispositions would also benefit more from instructions and practice than participants with lower scores on these dispositions). The study also applied measures of mental effort to explore the effects of instructions on cognitive load.

The experiment presented in Chapter 3 aimed to replicate and extend findings from the study in Chapter 2. This experiment was with first, third and fourth year continuing education students of the Economics and Business Education department of (*N* = 183). It was again investigated whether general critical thinking instruction would need to be combined with domain-specific practice in order to improve reasoning performance, but it was also investigated whether improved reasoning performance would still remain after a three-week delay. Again, it was investigated whether dispositions impacted reasoning performance prior to and after instruction.

The experiment presented in Chapter 4 aimed to address some issues that remained unclear in Chapter 2 and 3, seeking to unravel the effectiveness of general critical thinking instructions, domain-specific practice, and self-explanation during practice, on reasoning performance. Participants were first year students from an Economics and Business Education department (*N* = 152). Next to the effects of different instructions, it was again investigated whether dispositions impacted reasoning performance prior to and after instruction. This study also explored the relationship between reasoning performance, confidence, and invested mental effort prior to and after instructions.

In the final chapter (Chapter 5) the main findings are summarized and discussed in terms of their implications for educational research and practice, and considerations for future research are provided.

Contents

# 2 Improving Students' Critical Thinking: Empirical Support for Explicit Instructions Combined with Practice[1]

1 This chapter is accepted with minor revisions as Heijltjes, A., Van Gog, T., & Paas, F. Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology*.

This experiment investigated the impact of different types of critical thinking instructions and dispositions on bias in economics students' (*N* = 141) reasoning performance. The following conditions were compared: A) implicit instruction, B) implicit instruction with practice, C) implicit instruction with explicit instruction and practice, D) implicit instruction with explicit instruction, practice, and self-explanation prompts, E) implicit instruction with explicit instruction, practice, and activation prompts. Results showed that explicit instruction combined with practice is required to improve critical thinking (i.e., conditions AB < CDE) Prompting during practice had no added performance benefits. Participants' dispositions towards actively open-minded thinking predicted their pre-test and post-test scores, but did not interact with instruction condition, suggesting that receiving explicit instruction combined with practice was equally effective for all students.

Students in higher education frequently show poor critical thinking strategies (Carrithers, Ling, & Bean 2008; Flores, Matkin, Burbach, Quinn, & Harding, 2012; Harasym, Tsai, & Hemmati, 2008; Tsui 2002). Critical thinking is considered a powerful resource and important skill to survive and succeed in the complex 21st century world (Pellegrino & Hilton, 2012). It is an important requirement in many dynamical and rapidly changing professional environments such as medicine (Croskerry, 2003) and economics (Klebba & Hamilton, 2007; Smith, 2003). In these environments, a lack of critical thinking skill can result in biased reasoning and consequently in erroneous decisions with severe consequences. Therefore, education has a major role to improve students' critical thinking (Facione, 2009; Niu, Behar-Horenstein, & Garvan, 2013; Paul, 1990) and training of critical thinking skills is important to reduce biased reasoning in professions that rely on complex decision making (Evans, 2003; West, Toplak, & Stanovich, 2008). Moreover, improving critical thinking is important from an educational perspective. Acquisition of critical thinking skills has often been found to lead to better learning and transfer of trained tasks, because it results in integrated knowledge structures and more generalized knowledge (Helsdingen, Van Gog, & Van Merriënboer, 2011), facilitates students' evaluations of their own thinking according to specified standards (Celuch & Slama, 1998), and makes knowledge more accessible and usable (Billing, 2007).

Although there are good reasons for seeing critical thinking as a desirable outcome of education (Braun, 2004; Halpern, 1998; Pithers & Soden, 2000), it is rarely explicitly taught (Paul, 2005). This also applies, for instance, to business and economics education (Jones, 2007; Smith, 2003), where the prevailing view seems to be that immersion in business methods and strategies will also automatically lead to the development of critical thinking skills (Borg & Borg, 2001; Jones, 2007). However, there is little evidence that such skills develop spontaneously as a consequence of instruction in a discipline (Halpern, 1999). On the contrary, research indicates that critical thinking improves mainly when instructed explicitly (Abrami et al., 2008; Ennis, 1989; Ritchart & Perkins, 2005).

Another reason for the lack of explicit instruction of critical thinking, might lie in the difficulties that educators in economics (and other domains for that matter) face when trying to derive guidelines from exisiting research regarding when, where, and how to foster critical thinking throughout the curriculum (Smith, 2003). According to the meta-analysis by Abrami et al. (2008), existing studies on critical thinking instruction often lack a powerful empirical design and often fail to disentangle the impact of particular instructional conditions (see also McMillan, 1989; Ten Dam & Volman, 2004; Wolcott, Baril, Cunningham, Fordham, & St.Pierre, 2002). A recent meta-analysis by Niu et al. (2013) showed that although critical thinking teaching interventions seemed to be benefical, the magnitude of the average effect was small. Niu et al. (2013) raised the question of how the teaching of critical thinking can be improved to yield more satisfying results and they suggested to explore factors that influence the effectiveness of critical thinking instructions in more detail, for example by taking into account student characteristics. In sum, despite a lot of prior research on this topic, more powerful experimental support for the impact of different kinds of critical thinking instruction for different kinds of students is required in general (Ten Dam & Volman, 2004), and particularly in economics education (Wolcott et al., 2001).

The present experiment examines the effects of different instructional methods on economics students' critical thinking. The study focuses on an essential aspect of critical thinking, which plays a key role in decision-making in complex and high-risk situations that often occur in the field of economics: the avoidance of biased reasoning by means of controlled rational judgment and decision-making (Evans, 2003; Smith, 2003; Stanovich & Stanovich, 2010; Tversky & Kahneman, 1983; West et al., 2008). Biased reasoning seems inherent to human reasoning and decision-making, and economists are no exception; they too seem prone to violating the principles of rationality (Sanfey, Loewenstein, McClure, & Cohen, 2006). Before describing the research on critical thinking instructions and the rationale behind our study, we will first discuss the principles of critical thinking in more detail, especially in relation to biased reasoning and its underlying cognitive mechanisms.

## Critical Thinking

In order to support the instruction and assessment of critical thinking, a consensus definition of critical thinking has been developed by an expert Delphi Panel of the 'American Philosophical

Contents

Association' (APA; Facione, 1990). This definition is very comprehensive, encompassing a variety of viewpoints on critical thinking, and includes both thinking dispositions and cognitive skills. It characterizes critical thinking as a "purposeful, self-regulatory judgment, which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." (Facione, 1990, p. 2). Dispositions refer to 'habits of mind' such as "open-mindedness concerning divergent worldviews, flexibility in considering alternatives and opinions, and willingness to reconsider and revise views when change is warranted" (Facione, 1990, p. 25). According to the APA experts, cognitive skills refer – amongst others- to skills such as interpretation, evaluation, and inference which in turn has been subdivided into subskills. Inference, for instance, has the subskills 1) querying evidence to judge which information is needed to support premises, questions or issues and to determine strategies to acquire such information, 2) conjecturing alternatives in order to develop a variety of options, think about their potential consequences and prioritize the alternatives in light of the goal and 3) drawing conclusions in order to apply appropriate inference and reasoning strategies (Facione, 1990).

Besides this description of criticical thinking by the APA, many other definitions have been proposed (Ennis, 1985; Halpern, 1998; Kuhn, 1999; Paul, 1990; Siegel, 1980). For example critical thinking has been defined as "reflective and reasonable thinking that is focused on deciding what to believe or to do" (Ennis, 1985, p. 45), as "the use of cognitive skills or strategies that increase the probability of a desirable outcome" (Halpern, 1998 p. 450), or as "disciplined self-directed thinking which exemplifies the perfections of thinking appropriate to a particular mode or domain of thought" (Paul, 1990, p. 25). Most of these critical thinking definitions show much overlap, but differ on what they emphasize as key characteristics, although some of them differ on whether critical thinking skills are viewed as domain general or domain specific, or about the most important components of the critical thinking concept. For example Ennis (1985) and Paul (1985) argued that critical thinking is domain general, that is independent of specific disciplines, whereas McPeck (1990) argued that criteria for applying and assessing critical thinking have to be derived from specific topics, subjects, fields or domains (i.e., domain specific). This distinction implies different views on how to instruct critical thinking. For instance, according to Ennis (1989) general critical thinking principles can be taught that are then applied to a specific domain, whereas McPeck (1990) argues that critical thinking cannot be instructed independently from a specific domain.

Taking a domain-general view, a number of researchers have linked critical thinking more strongly to the concept of rationality (Siegel, 1989; Stanovich, 2011; West et al., 2008), which is also the focus of our study. Siegel argues for instance, that education should aim to foster rationality by focusing on reasons on which assessments, judgments and actions are based. By identifying thinking errors such as reasoning bias, the degree of an individual's rationality can be assessed (Gilovich & Griffin, 2002; Tversky & Kahneman, 1981). According to West et al. (2008), studying biased reasoning and how to prevent it is an important area of research that complements traditional research on critical thinking. For instance, the literature on critical thinking in educational research has focused on a wide variety of topics in the past decades (Angeli & Valanides, 2009), but has paid little attention to the aspect of the avoidance of biased reasoning (West et al., 2008).

Biased reasoning has been demonstrated on a range of reasoning tasks. For example, entrepeneurs have been shown to neglect relevant base-rates and to be overly confident in starting a firm and doing business (Cooper, Woo, & Dunkelberg, 1988; Kahneman & Lavallo, 1993). Moreover, many people tend to accept a conclusion from premises if it is believable regardless of its logical validity (belief bias effects on syllogistic reasoning tasks; e.g., Evans, Handley, & Harper, 2001), tend to consider only information which matches the lexical content of a statement about which they are reasoning as being relevant, and conversely, tend to neglect logically relevant information (matching bias on selection tasks; e.g., Evans, 2003), and tend to show erroneous judgments and reasoning on causal and probabilistic tasks, assessment of covariation tasks, framing tasks, and conjunction tasks (Gilovich, Griffin, & Kahneman, 2002; Stanovich, 2011; Tversky & Kahneman, 1983).

Biased reasoning can be explained by dual processing models (Evans, 2003, 2008, 2011; Stanovich, 2011), which are based on the assumption that two types of cognitive processes are involved in reasoning. Type 1 reasoning has a rapid, automatic nature, involves little reflection and imposes a relative low load on working memory, which allows for rapid processing of large amounts of information. Decision-making using Type 1 reasoning is based on past experiences, which is useful and efficient in many routine situations. However, because of its automaticity, it might result in biased thinking in other, non-routine situations. Type 2 reasoning can overrule these automatic responses by explicit reasoning efforts. Type 2 reasoning is slow and sequential in nature, requires the exclusion of attention to other matters, and is associated with rational thinking, and as such, it imposes a higher load on working memory than Type 1 reasoning (Evans, 2011).

Type 2 reasoning seems to be associated with thinking dispositions (Stanovich, 2011), such as 'actively open-minded thinking' (i.e., AOT: the active search for evidence against one's own beliefs, plans, or goals and the ability to weigh available evidence fairly; Baron, 2008) and 'need for cognition' (NFC, i.e., "the tendency for an individual to engage in and enjoy thinking"; Cacioppo, Petty, & Kao, 1984, p. 306). Research has shown that AOT and NFC predicted performance on several of the reasoning tasks mentioned above (West et al., 2008). For instance, individuals with higher AOT scores were found to perform better on argument evaluation (Stanovich & West, 1997) and avoiding unsophisticated responses in argument (Sá, Kelly, Ho, & Stanovich, 2005). According to Baron (2008), individuals

Contents

with higher AOT scores considered more alternative possibilities (i.e., other possibilities than initially favored), asked themselves more frequently what possibility would produce the best expected outcomes, weighed available evidence more fairly, and showed less overconfidence in hasty conclusions than individuals with lower AOT scores. NFC is characterized by cognitive motivation, and individuals with higher NFC scores have, for example, been found to recall more of the information to which they are exposed, to be influenced more by the quality of arguments, to generate more task-relevant thoughts, and to put more effort into cognitive tasks than individuals with lower NFC scores (for a review, see Cacioppo, Petty, Feinstein, & Jarvis, 1996). Thus, it can be expected that participants with higher scores on AOT and NFC would score better on a (initial) reasoning test, but it is an open question whether they would also benefit more from critical thinking instructions than participants with lower scores on these dispositions.

When Type 2 reasoning fails to override Type 1 reasoning, this can lead to biases and fallacies in judgments and decisions (Stanovich, 2009). Such failures to override Type 1 reasoning might be due to a lack of declarative knowledge about, for example, probability or conjunction rules, or, when such knowledge is present, due to insufficiently developed strategies such as hypothetical thinking (i.e., engaging in cognitive simulations to consider alternative responses or possibilities by decoupling secondary representations from primary representations; Stanovich, 2009). Both a lack of declarative knowledge and knowledge of reasoning strategies might presumably be counteracted by explicit instructions. Because Type 2 reasoning is more effortful than Type 1 reasoning (Evans, 2011), it is possible that such instructions would lead to more investment of mental effort on a subsequent test. On the other hand, knowledge of strategies gained through instructions might reduce the cognitive load imposed by the task and therefore lead to less effort investment on a subsequent test.

## Critical Thinking Instruction to Avoid Biased Reasoning

According to Larrick (2004), instead of aiming to identify and demonstrate biases and thinking fallacies, researchers should devote more attention to finding effective debiasing techniques and strategies (see also Klaczynski, Gordon, & Fauth, 1997). In contrast to critical thinking instruction in general (for a meta-analysis see Abrami, 2008), relatively few studies have focussed on the specific aspect of instructing how to avoid biased reasoning, that is, on the question of whether rational thinking is trainable, and if so, what type of instruction is most beneficial (Ritchart & Perkins, 2005; Stanovich, 2011). The available research, however, reveals some evidence for positive effects of explicit training and raises the question of whether that should be combined with opportunities for practice.

For example, Larrick, Morgan, and Nisbett (1990) found that participants who received instructions on cost-benefit principles became more rational in reasoning (e.g., less prone to the economic trap of considering past investment in terms of money, time, or effort, in making current decisions) than participants who received no instructions. Macpherson and Stanovich (2007) showed that participants who were instructed to decouple prior beliefs and opinions (i.e., myside bias), from evaluation of evidence and arguments improved their performance on argument generation tasks. The effect of instructions was strongest when participants received quite explicit instructions regarding unbiased responding (i.e., prompting them with 'give both reasons for and reasons against') and it was weakest and not significant when participants only had to indicate what they should do to avoid bias. Fong, Krantz, and Nisbett (1986) found that explicitly teaching reasoning principles on reasoning tasks such as instructing the law of large numbers enhanced performance on base-rate problems (i.e., problems in which one has to take sufficient account of the prior probability information) and employing statistical concepts in analyzing everyday inferential problems. They demonstrated that both formal rule training (i.e., on the law of large numbers) and guidance to apply these rules in particular domains by means of examples, were effective for improving reasoning performance. The combination of explicit rule training plus example training improved the reasoning process and had an additive effect (as high as the sum of the effects of each type of training in isolation).

However, on selection problems (e.g., presenting a conditional rule, *if p then q*, and testing the truth or falsity of that rule by selecting the right options, *p* and *not-q*) Cheng, Holyoak, Nisbett, and Oliver (1986) found that neither rule training by itself nor example training by itself improved reasoning, but training on rules coupled with example training improved reasoning performances significantly. They argued that rule training only is ineffective, as people lack the ability to apply abstract logical rules to concrete problems, whereas example training in isolation fails because people have no intuitive version of the rules that have to be applied.

The literature on teaching critical thinking in general, has shown that explicit critical thinking instruction combined with the opportunity to integrate the principles that were taught with subject matter, works best (e.g., Abrami et al., 2008; Angeli & Valanides, 2009; Bangert-Drowns & Bankert, 1990). Recently, Marin and Halpern (2011) compared the effectiveness of imbedded critical thinking instruction (i.e., instructions woven into the content matter) with explicit instruction (i.e., specific supplementary critical thinking lessons). Participants who received explicit instruction showed larger gains on critical thinking than participants in the imbedded group and moreover, they transferred those skills to everyday situations. Bransford, Sherwood, Vye, and Rieser (1986) argued that explicit thinking instructions alone are not sufficient; they should

Contents

be combined with practice in a domain-specific context, because general strategies and specific knowledge both are needed to allow students to integrate general rules, and to combine and cluster information in a meaningful way. Such well-organized knowledge will increase the probability that relevant information will be accessed when needed.

In sum, these findings suggest that different instructional methods might not be equally effective for improving unbiased reasoning, as it might benefit from explicit instruction (Fong et al., 1986; Macpherson & Stanovich, 2007), example training (Fong et al.,1986), or the combination of both (Cheng et al., 1986). Critical thinking in general –not explicitly the avoidance of biased teasoning- seems to benefit from explicit instruction (Bangert-Drowns & Bankert, 1990) combined with integration to subject matter (Abrami et al., 2008). Based on this literature review it can be expected that explicit critical thinking instructions are necessary to improve reasoning performance, but that they might have to be combined with practice in order to be (more) effective.

Assuming that both explicit instruction and opportunities for applying the reasoning principles that were instructed, would be required for effective acquisition of reasoning skills, then the question arises whether prompts during applying reasoning principles might have added value and if so, what kind of prompts?

One potentially effective strategy is prompting self-explanations. Austin, Gregory and Chiu (2008) showed that on base-rate, conjunction, and gambler fallacy tasks, pharmacy students who were prompted to provide just a brief written explanation and justification of how they arrived at their answer, demonstrated a significantly better performance on critical thinking, than students who did not receive these prompts. This finding is consistent with studies about the effectiveness of self-explanation. According to Roy and Chi (2005), self-explanation fosters the proper use of available knowledge and skills. The beneficial effects of self-explanations occur even though participants rarely receive feedback on the quality of their explanations (Matthews & Rittle-Johnson, 2009). It is assumed that self-explaining information can promote learning of reasoning tasks through the integration of new information with prior beliefs (Lombrozo, 2006). For instance, self-explaining will help people to assess the probability of claims in light of their prior beliefs and might evoke beliefs-revision (Lombrozo, 2006; Williams & Lombrozo, 2010). As an essential aspect of unbiased (Type 2) reasoning is to decouple prior beliefs from evidence, and prompting self-explanation during practice with reasoning tasks might have an additional impact on acquisition of reasoning skills above and beyond explicit instructions combined with practice without such prompts. Therefore it can be expected that providing self-explanation prompts during practice would have an additional beneficial effect on the acquisition of reasoning skills.

Another potentially effective strategy to reduce bias might be the use of prompts that draw attention away from salient but non-relevant information and instead activate relevant but less salient information through the use of activation prompts. Moutier, Angeard, and Houdé (2002) showed that inhibition training improved performance on Wason selection tasks (Wason, 1968) and Moutier and Houdé (2003) showed that inhibition training improved participants' reasoning performance on classical conjunction tasks (Tversky & Kahneman, 1983). For example, on conjunction tasks, participants were taught to inhibit the misleading scheme (the conjunction) and to activate the relevant scheme (the probability). Participants' performance increased when they were encouraged to redirect their attention towards logically relevant information. The underlying assumption is that reasoning biases are caused by an inhibition failure in working memory and not due to a lack of understanding of probabilistic principles or a lack of deductive competence. Stanovich and Stanovich (2010), however, stated that suppressing the initial response (Type 1 reasoning) is only helpful when a better response is available to substitute for it. Therefore, the use of prompts to activate the 'weaker' cues instead of prompts to suppress dominant cues might be more effective (i.e., evoke an attention shift from irrelevant to relevant information; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Also, it would be reasonable to expect that such 'activation prompts' during practice would encourage an attention shift to relevant tasks aspects, which might have an additional impact on the acquisition of reasoning skills above and beyond explicit instructions combined with practice without such prompts.

In sum, an important question for education, which is addressed in the present study, is whether critical thinking in terms of unbiased reasoning can be enhanced by explicit instructions in combination with practice, and whether self-explanation or activation prompts during practice can further enhance reasoning skills.

## The Present Study

The present study addresses the question of whether economics students' critical thinking skills, defined as their performance on reasoning tasks that are prone to biased reasoning, will improve through implicit critical thinking instructions provided in a regular course context, or whether explicit instructions are needed and if so, whether these need to be combined with practice. In addition, it is investigated whether self-explanation and activation prompts during practice have added value for later reasoning performance. Finally, this study also explores the role of students' dispositions in (acquiring) reasoning skills, and applies measures of mental effort to explore the effects of instructions on cognitive load.

Regarding dispositions, it is hypothesized that participants with higher scores on dispositions (i.e., AOT and NFC) would score better on the pre-test; whether they would

Contents

also benefit more from instructions and practice than participants with lower scores on these dispositions is an open question that we will explore.

Regarding the effects of instructions, it is hypothesized that explicit critical thinking instructions are necessary. If so, then the conditions receiving not only implicit instructions through the regular course, but also explicit instructions would perform better on the post-test than the conditions receiving only implicit instructions. Moreover, if explicit critical thinking instructions need to be combined with practice in order to be effective, then post-test performance in the explicit instructions conditions would only improve compared to the other conditions on the task categories that were practiced after instruction, but not on task categories that were instructed but not practiced. If practice only is sufficient, then the implicit instructions plus practice condition would outperform the implicit instructions only condition on the practiced task categories. If providing self-explanation or activation prompts during practice would have an additional impact on the acquisition of reasoning skills, then the post-test performance of the prompting conditions would be higher compared to the condition that received explicit instructions without prompts.

Finally, potential differences between conditions in invested mental effort are explored, as these might reveal effects of instructions and practice on the cognitive demands imposed by reasoning tasks. On the one hand, educational research has shown that knowledge of strategies gained through instructions might reduce the cognitive load imposed by the task and therefore lead to less effort investment on a subsequent test (Paas, Renkl, & Sweller, 2003). On the other hand, Type 2 reasoning is more effortful than Type 1 reasoning (Evans, 2011), so on reasoning tasks, instructions might lead to more effort investment on a subsequent test.

# Method

## Participants and Design

Participants were 141 second-year students (57 male; age *M* = 20.81, *SD* = 1.57) from an Economics and Business Education department of a Dutch University of Applied Sciences. Participants were randomly assigned to one of five conditions: A) implicit instruction (*n* = 25), B) implicit instruction with practice of some of the task categories (*n* = 28), C) implicit instruction with explicit instruction of all task categories, and practice of some of the instructed task categories (*n* = 31), D) implicit instruction with explicit instruction of all task categories, and practice of some of the instructed task categories, and self-explanation prompts during practice (*n* = 27), E) implicit instruction with explicit instruction of all task categories, and practice of some of the instructed task categories, and activation prompts during practice ($n$ = 30). Table 1 shows an overview of the experimental design.

## Materials

**Critical thinking disposition tests**. We used a Dutch translation of the 41-item Actively Open-minded Thinking test (AOT; Stanovich & West, 2007) and the 18-item (short form) Need for Cognition Scale (NFC; Cacioppo et al., 1984) to measure critical thinking dispositions. Both consist of items requiring a response on a 6-point Likert scale (AOT: as in Stanovich & West, 2007; NFC: as in West et al., 2008; note that Cacioppo et al., 1984 used a 9-point scale) ranging from (1) strongly disagree to (6) strongly agree. Scores on the items are summed for AOT and for NFC separately (after reverse scoring items that are formulated negatively). Higher scores on AOT characterize a greater tendency toward open-minded thinking. The reliability of the AOT was good: the Cronbach's alpha was .84. Higher scores on the NFC represent a greater tendency to engage in and enjoy effortful cognitive endeavors. Reliability of the NFC was also good: the Cronbach's alpha was .78.

**Critical thinking skills tests**. The critical thinking skills tests consisted of sixteen tasks. The tasks in the pre-test and post-test were structurally equivalent, but surface features (cover stories) differed. The reasoning tasks examined the heuristic tendency: 1) to be influenced by intense personal and case evidence in favor of more representative statistical evidence (two causal base-rate tasks adapted from Fong et al., 1986), 2) to base judgments on prior belief and intuition without taking sufficient account of the prior probability (two non-causal base-rate tasks adapted from De Neys & Glumicic, 2008), 3) to neglect simple and fundamental qualitative rules of probability in conjunction problems in which a conjunction cannot be more probable than one of its ingredients (two conjunction tasks, adapted from Tversky & Kahneman, 1983), 4) to shift outcomes when the same information is framed in different ways as in cases of gains and risky options (two framing tasks adapted from Tversky & Kahneman, 1981), 5) to evaluate the information given in a 2 x 2 contingency table unequally, that is, to base estimations on already experienced evidence and disregard some of the presented evidence (two covariation tasks adapted from Wasserman, Dorner, & Kao, 1990), 6) to examine the tendency to verify rules rather than to falsify them (two Wason selection tasks adapted from Stanovich, 2009, and Wason & Shapiro, 1971), and 7) to examine the tendency to evaluate the logical validity of arguments on the basis of one's prior beliefs about the truth of conclusions on syllogistic reasoning tasks (Evans, 2003; Markovits & Nantel, 1989; Sá et al., 1999). In syllogistic reasoning tasks a conclusion is drawn from two given premises or assumed propositions and participants have to indicate whether

Contents

**Table 1**

*Overview of the Study Design with Sequence of Events and Time (min.) Devoted to Instructions.*

| | | | | | |
|---|---|---|---|---|---|
| *Week 1* | *Pre-test reasoning skills and dispositions (N = 141)* ➔ | | | | |
| *Week 2 to 6* | *Implicit instruction during course: 3 training sessions with fellow students and teacher, studying materials individually, and 2 assignments* ➔ | | | | |
| *Week 7* | *Group A (n = 25)* | *Group B (n = 28) Unrelated video (15 min)* ➔ | *Group C (n = 31) Video Critical Thinking Instruction (15 min)* ➔ | *Group D (n = 27) Video Critical Thinking Instruction (15 min)* ➔ | *Group E (n = 30) Video Critical Thinking Instruction (15min)* ➔ |
| | | *Practice 4 (out of 7) categories of tasks on a business case* | *Practice 4 (out of 7) categories of tasks on a business case* | *Practice 4 (out of 7) categories of tasks on a business case + self-explanation prompts* | *Practice 4 (out of 7) categories of tasks on a business case + activation prompts* |
| | *(-)* ➔ | *(±20 min)* ➔ | *(±20 min)* ➔ | *(±20 min)* ➔ | *(±20 min)* ➔ |
| | *Post-test reasoning skills* | | | | |

*Note.* Group A: implicit instruction only, Group B: implicit instruction + practice, Group C: implicit instruction + explicit instruction + practice, Group D: implicit instruction + explicit instruction + practice + self-explanation prompts, Group E: implicit instruction + explicit instruction + practice + activation prompts.

or not the conclusion follows logically from the premises. Different types of inconsistent (i.e., the validity of the conclusion is in conflict with its believability) syllogistic tasks were included (one of each type): affirming the antecedent or modus ponens (if p then q, p therefore q; valid), affirmation of the consequent (if p then q, q therefore p; invalid), denial of the antecedent (if p then q, not p therefore not q; invalid) and denial of the consequent, or modus tollens (if p then q, not q therefore not p; valid). An example of an inconsistent syllogistic reasoning task (affirmation of the consequent) modeled after Stanovich et al. (2008) is:

*Please indicate whether the conclusion follows logically from the premises.*

*Premise 1. All oil companies are quoted companies*

*Premise 2. Shell is a quoted company.*

*Conclusion: Shell is an oil company.*

*Answer option A: This conclusion follows logically from the premises.*

*Answer option B: This conclusion does not follow logically from the premises*.

Many people will choose Answer A, however, this is not correct because the line of reasoning 'if p then q, q therefore p' (i.e., affirmation of the consequent), is invalid.

The content of the surface features (cover stories) of the tasks was adapted to the interests of students in the economics domain. The format of the tasks differed; a multiple-choice format with two to five answer options (depending on task characteristics) was used (the correct answer based upon rational reasoning strategies and incorrect answers related to biased reasoning). The internal consistency (Cronbach's alpha) of the two tests was .49 and .66 respectively. Modest reliability for the composite score could be expected and is in line with previous research with this type of tasks (de Bruin, Parker, & Fischhoff, 2007; West et al., 2008). Also, it should be noted that reliability on the post-test was higher, which is logical as performance on a pre-test prior to instruction is more random or variable for these kinds of tasks than performance on a post-test after instruction.

**Mental effort.** Invested mental effort was measured with a 9-point subjective rating scale ranging from (1) very, very low effort to (9) very, very high effort (Paas, 1992; Paas & Van Merriënboer, 1993). Mental effort is an indicator of actual cognitive load experienced, and this scale is widely used in educational research (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Van Gog & Paas, 2008).

**Critical thinking instructions**. The implicit instruction (which all conditions received) consisted of a seven-week regular economics course about argumentation and negotiation skills that –amongst others- included three training sessions and two assignments. In advance of the training sessions students studied written

Contents

materials about negotiation and resistance to manipulative techniques (e.g., asking for facts and figures from people who might bluff, keeping the public interest in mind instead of relying on personal concerns, becoming more sceptical towards unreliable authority) and argumentation skills (e.g., substantiating arguments, recognizing types of arguments, and argumentation errors). The assignment consisted of practicing the negotiation and argumentation skills on a business case and plenary presenting a critical reflection on the case to the teacher and fellow students. Thus, in the implicit instruction condition general critical thinking and reasoning skills are instructed and practiced on a business case, but no explicit instructions are provided concerning the rules that apply specifically to the reasoning tasks used in this study.

Participants in the condition that received implicit instruction + practice of some task categories (B) also took the same course, but were additionally provided with a video-based presentation on an unrelated topic (i.e., a 15 min. digital video on what happens in your brain when you are in love). Then participants were exposed to a business-case from an economics course containing a description of a successful coffee manufacturer who had to decide about marketing, quality control, extending the assortment, and the sustainability measures. Four categories of tasks (i.e., a contingency task, a conjunction task, a non-causal base-rate task, and a Wason selection task) were practiced with a similar format as the tasks used in the pre-tests, but the cover stories of the practice tasks were derived from the business case. Thus, the reasoning principles had to be applied to an economics context. See Appendix A for an overview of the practiced tasks. Participants did not receive feedback on the quality of their performance on the practice tasks.

Participants in the condition that received implicit instruction + explicit instruction on all task categories + practice of some task categories (C) took the same course as condition A and B, but were additionally provided with explicit instructions concerning the reasoning tasks used in this study, by means of a 15 min. computer-based video presentation in which general information about critical thinking was provided, such as the features of critical thinking and its importance, the required reasoning skills, the dispositions, and the risk of biased thinking and fallacies in thinking. Then characteristics of a critical thinker were presented (as verbal instructions with visual support on the slides consisting of key words and images) coupled with examples and demonstrations of all task categories, referring to the tasks seen in the pre-test, which allowed participants to mentally correct initially erroneous responses. As an example, participants received the auditory information *"Critical thinking requires many skills; for example inductive and deductive reasoning skills… Inductive reasoning deals with the question how to obtain a conclusion from specific information. The way from specific information to a general conclusion is one with pitfalls… Before positing a statement a critical thinker will search for information not only based on individual experiences but* *will also search for information from reliable resources*" after which the base-rate fallacy was demonstrated by a slide presenting an image of a person with a guitar (the image was such that the person could not be identified as male or female), a picture of beer, and a picture of some engineering tools. Next to the image the written text appears: base-rate fallacy, group: 990 women, 10 men. At the same time the following auditory information was given:*"The base-rate fallacy is a thinking failure that occurs when the statistical distribution of a population is ignored. For example when we select a particpant randomly out of 1000 participants with 990 women and 10 men, and we will tell you that this person is called Sam, loves drinking beer and listening to hardrock music, and has graduated as mechanical engineer, then most people tend to find it most likely that this person is a man. In that case they ignore that the total group contains 10 men only"* (NB: a different name was used in the video, a Dutch name that can also be given to both men and women, but for the sake of clarity we used an English name here). Another example: the Wason selection task was presented, consisting of a conditional statement in written text and an image of four boxes. Each box had a letter on one side and a number on the other side (two showed the letter and two showed the number). Participants had to decide which two boxes should be turned over to find out whether the statement was true or false. With this slide, the following verbal information was given: *"Applying abstract logical rules may be difficult. For example on the front side of these boxes there is a letter X or Y and on the other side there is a number 1 or 2. The rule is 'if there is an X on one side of the box, then number 1 is on the other side of the box'. The question is: which two boxes need to be turned over to decide whether the rule applies or not? You probably tend to turn over box X and 1, but will that lead you to the right answer?"*. Then participants received visual and auditory information about all possibilities (i.e., what will happen if selecting X, what will happen if selecting Y, etc.), in a step-by-step demonstration, checking all possible -right and wrong- answers. Similar procedures were used for the other five task categories. The computer controlled the sequence and pace of the video, but participants could pause, forward, and rewind. After the video-instruction all participants were exposed to the same business-case as in condition B.

Participants in the implicit instruction + explicit instruction on all task categories + practice of some task categories + self-explanation condition (D) were exposed to the same procedure as participants in condition C but were additionally given prompts after each of the four practiced task categories to self-explain how the answer was obtained: 'Explain by using keywords how you've come to the answer'. Participants in the implicit + explicit instruction on all task categories + practice of some task categories + activation condition (E), differed only from condition D with regard to the prompts given, that is, the prompts attended participants to focus on the relevant factors of the four practiced task categories. Participants were prompted with hints: 'search for

Contents

confirmation and refutation' on the contingency task, 'think of the logical probability' on the conjunction task, 'think of the statistical distribution' on the non-causal base-rate task, and 'violation?' on the Wason selection task.

### Procedure

At the beginning of the seven-week economics course, all participants completed the computer-based pre-test and the disposition tests. Then all participants followed the regular training sessions of the course. In week 7 participants in conditions B, C, D and E had an additional session of approximately 35 min. duration at the institute, during which the participants in condition C, D and E received the critical thinking instruction and condition B received the unrelated video, followed by practice of 4 categories of reasoning tasks framed in the context of a business case in these conditions. Immediately after this session attended only by participants in conditions B, C, D and E, the post-test session took place for participants in all five conditions. Invested mental effort was rated after each task on the pre-test and the post-test and in conditions B, C, D and E also after each practice task.

### Data analysis

For each correct answer on the critical thinking skills pre-test and post-test, 1 point was assigned, resulting in a maximum score of 8 points for practiced task categories and 8 points for non-practiced task categories. Dispositions scores were obtained by summing the rating on AOT after reverse scoring negatively formulated items, resulting in a maximum score of 246. For all analyses, a significance level of .05 was used.

## Results

An ANOVA confirmed that random assignment of participants to conditions had been succesful; there were no significant diffences among conditions in critical thinking pre-test performance on practiced tasks, $F(4, 136) = .31$, $p = .870$, $\eta_p^2 = .01$, not-practiced tasks, $F(4, 136) = 1.82$, $p = .129$, $\eta_p^2 = .05$, mental effort invested in the pre-test on reasoning tasks, $F(4, 136) = .42$, $p = .793$, $\eta_p^2 = .01$, critical thinking dispositions: AOT, $F(4, 136) = .54$, $p = .709$, $\eta_p^2 = .02$, NFC, $F(4, 136) = .38$, $p = .822$, $\eta_p^2 = .01$, or educational background, $\chi^2(12) = 8.82$, $p = .718$.

Contents

## Dispositions

To investigate the hypothesis that participants with higher scores on dispositions (i.e., AOT and NFC) would score better on an initial assessment of critical thinking skills than participants with lower scores on these dispositions, a regression analysis was used. Overall dispositions scores significantly predicted pre-test critical thinking skills, $R^2 = .09$, $F(2, 138) = 6.56$, $p = .002$; however, only AOT significantly predicted pre-test performance, $\beta = .31$, $t(138) = 3.47$, $p = .001$, but NFC did not, $\beta = -.04$, $t(138) = -.42$ $p = .676$. Multiple regression analysis were conducted to explore whether AOT and NFC would predict post-test performance after controlling for pre-test scores. Again, dispositions scores significantly predicted post-test critical thinking skills, $\Delta R^2 = .06$, $F_{change}(2, 137) = 5.08$, $p = .007$, $R^2 = .20$; however, whereas AOT significantly predicted post-test performances positively, $\beta = .25$, $t(137) = 2.93$, $p = .004$, NFC did so negatively, $\beta = -.19$, $t(137) = -2.26$ $p = .025$. To investigate whether dispositions interacted with instructions on post-test performance a multiple regression analysis, corrected for pre-test-critical thinking skills, revealed no significant interaction between AOT scores and instructions conditions, $\Delta R^2 = .02$, $F_{change}(4, 133) = .86$, $p = .489$, or between NFC scores and instruction conditions $\Delta R^2 = .01$, $F_{change}(4, 129) = .45$, $p = .769$.

## Critical Thinking Test Performance

Table 2 provides the means and standard deviations of critical thinking skill scores in each condition on the pre-test and the post-test and the adjusted means and standard deviations on the post-test. A repeated measures ANCOVA with practice (practiced vs. not-practiced tasks) as within-subjects factor, instruction condition as between-subjects factor, and overall pre-test score as covariate, showed a significant main effect of instruction condition, $F(4, 135) = 6.54$, $p < .001$, $\eta_p^2 = .16$, no significant main effect of practice, $F(1, 135) = 1.17$, $p = .281$, $\eta_p^2 = .01$, and an interaction effect of instruction condition with practice, $F(4, 135) = 13.82$, $p < .001$, $\eta_p^2 = .29$. To investigate this interaction, two separate ANCOVAs were conducted on practiced and not-practiced task, with overall pre-test score as covariate. This revealed a significant effect of instruction condition on post-test performance on practiced tasks, $F(4, 135) = 12.43$, $p < .001$, $\eta_p^2 = .27$, and a significant effect of condition on post-test performance on not-practiced tasks $F(4, 135) = 2.58$, $p = .040$, $\eta_p^2 = .07$. On practiced tasks, Bonferroni post-hoc tests revealed significantly higher performance in the conditions C (implicit plus explicit instruction of all task categories plus practice of some task categories), D (implicit plus explicit instruction of all task categories plus practice of some task categories plus self-explanation prompts), and E (implicit plus explicit instruction of all task categories plus practice of some task categories plus activation prompts) compared to condition

Contents

**Table 2**

*Means and Standard Deviation of Pre-test and Post-test Scores on Practiced (max. 8) and Not-practiced Tasks (max. 8)*

| Variables | | Pre-test | | Post-test | | Adjusted Post-test* | |
|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *M* | *SD* | *M* | *SE* |
| Condition A: Implicit instruction | 25 | | | | | | |
| Practiced tasks | | 3.64 | 1.38 | 2.88 | 1.56 | 2.80 | .35 |
| Not-practiced tasks | | 4.60 | 1.29 | 4.76 | 1.30 | 4.68 | .25 |
| Condition B: Implicit instruction + practice | 28 | | | | | | |
| Practiced tasks | | 3.43 | 1.00 | 2.82 | 1.66 | 2.81 | .33 |
| Not-practiced tasks | | 3.54 | 1.67 | 4.82 | 1.09 | 4.81 | .24 |
| Condition C: Implicit instruction + explicit instruction + practice | 31 | | | | | | |
| Practiced tasks | | 3.42 | 1.36 | 5.10 | 1.87 | 5.28 | .31 |
| Not-practiced tasks | | 3.81 | 1.54 | 4.71 | 1.53 | 4.89 | .23 |
| Condition D: Implicit instruction + explicit instruction + practice + self-explanation prompts | 27 | | | | | | |
| Practiced tasks | | 3.41 | 1.60 | 4.63 | 2.06 | 4.66 | .33 |
| Not-practiced tasks | | 4.41 | 1.28 | 3.96 | 1.60 | 3.99 | .24 |
| Condition E: Implicit instruction + explicit instruction + practice + activation prompts | 30 | | | | | | |
| Practiced tasks | | 3.70 | 1.29 | 4.80 | 1.86 | 4.66 | .32 |
| Not-practiced tasks | | 4.77 | 1.65 | 5.07 | 1.34 | 4.93 | .23 |

Note. * Adjusted for pre-test scores

B (implicit instruction and practice of some task categories) and condition A (implicit instruction only), all *ps* $< .002$. No other comparisons were significant. On not- practiced tasks, Bonferroni post-hoc tests revealed no significant differences between conditions, all *ps* $> .064$.

### Mental Effort

In the practice phase, there were no significant diffences between conditions B, C, D and E (condition A did not practice tasks) in invested mental effort, $F(3, 112) = .37$, $p = .776$, $\eta_p^2 = .01$. To investigate the effects of instructions and practice on the cognitive demands of reasoning tasks a repeated measures ANOVA with invested mental effort on practiced tasks (invested mental effort in practiced vs. not-practiced tasks) and test time (invested mental effort in pre-test versus invested mental effort in post-test) as within-subjects factors and instruction condition as between-subjects factor, showed no significant effect of instruction conditions on invested mental effort, $F(4, 136) = .57$, $p = .687$, $\eta_p^2 = .02$, nor any interaction effects between instruction condition and invested mental effort on practiced versus not-practiced tasks, $F(4, 136) = .23$, $p = .923$, $\eta_p^2 =$ 01 or between instruction condition and invested mental effort on practiced versus not-practiced tasks and test time, $F(4, 136) = 1.04$, $p = .390$, $\eta_p^2 = .03$. Significant main effects were found for test-time, $F(1, 136) = 10.06$, $p = .002$, $\eta_p^2 = .07$, indicating that invested mental effort was higher on the pre-test ($M = 3.43$, $SD = 1.02$) than on the post-test ($M = 3.22$, $SD = .96$; $t(140) = 3.18$, $p = .002$, Cohen's $d = 0.21$), and a significant main effect was found of practiced versus not-practiced tasks, $F(1, 136) = 277.55$, $p < .001$, $\eta_p^2 = .67$, indicating higher invested mental effort on practiced tasks ($M = 3.77$, $SD = 1.03$) than on not-practiced tasks ($M = 2.94$, $SD = .88$; $t(140) = 16.87$, $p < .001$, Cohen's $d = 0.87$).

## Discussion

The present study supports our hypothesis that in order to enhance critical thinking skills in terms of performance on reasoning tasks that are prone to bias, the attention paid to critical thinking within a regular economics course was not sufficient. Instead, explicit critical thinking instruction combined with practice opportunities was necessary to improve performance; explicit instructions only without practice opportunities did not improve performance compared to no explicit instructions, and practice without explicit instructions did not improve performance compared to no practice. Hence, learning unbiased reasoning seems to be a function of two components: acquiring elementary competencies to perform specific cognitive reasoning operations from instructions and opportunities to practice these competencies in an economics

Contents

context (i.e., on a business case). This result expands findings from studies on teaching other aspects of critical thinking, which showed that explicit instructions improved critical thinking (e.g., Abrami et al., 2008; Angeli & Valanides, 2009). It is also in line with findings that rational thinking (Type 2 reasoning) to override Type 1 reasoning processes is trainable by explicit instruction (e.g., Fong et al., 1987), not only on base-rate and selection tasks but also on conjunction and co-variation tasks for students in the domain of higher economics education. The fact that performance was only better on the tasks that were explicitly instructed *and* practiced suggests that awareness of biased thinking alone by explicit instruction was not sufficient to recognize and apply the principles of unbiased reasoning (Larrick, 2004), and that the use of the instructions during practice in a domain specific context is important for improving performance. Possibly, the opportunity to apply instructions would enable participants to 'organize their knowledge' which would facilitate retrieving it from memory in subsequent reasoning tasks (Bransford et al., 1986).

Combining practice with prompts did not lead to better performance. Neither self-explanation prompts, which we expected to foster the proper use of available knowledge (Roy & Chi, 2005), nor activation prompts, which we expected to be able to affect critical thinking performance by redirecting attention to relevant cues (e.g., Moutier & Houdé, 2003), enhanced performance on practiced tasks compared to the condition that received explicit instructions and practice without prompts. Possibly, learning reasoning principles by explicit declarative instruction and then applying them by practicing tasks allows learners sufficiently to construct their understanding and evoke an attention shift to relevant task cues and inhibit initial automatic responses during practice. When providing instructions with examples, as was done in the explicit instruction, the way in which learners' actively process these examples by means of follow-up activities affects learning (Atkinson & Renkl, 2007; Lee & Anderson, 2013). Possibly, practice after instructions was an appropriate and sufficient way to actively focusing participants' attention to the central concepts and principles of the reasoning tasks.

Invested mental effort was investigated to explore potential effects of instructions and practice on the cognitive demands imposed by reasoning tasks. Two options were plausible: knowledge of strategies gained through instructions might reduce cognitive load imposed by the task, and therefore lead to less effort investment on the post-test (Paas et al., 2003), or instructions might evoke engaging in more effortful reasoning on these type of tasks (Evans, 2011). We found no difference in invested mental effort scores across all conditions and tests whereas on post-test scores, the implicit instructed plus explicit instructed plus practice condition (condition C) and the two additionally prompted conditions (conditions D and E), outperformed the other conditions (A and B), This seems to indicate an instructional effect on efficiency of performance in conditions C, D and E: a better performance was reached with equal investment of effort during practice and after instruction (for a discussion of efficiency, see Hoffman & Schraw, 2010; Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). The results on invested mental effort suggest that instructed participants did not engage in more effortful reasoning, that according to Evans (2011) could be expected, but allocated their cognitive resources more efficiently; however it is not clear whether effort investment to measure instructional efficiency can be taken as a measure of reasoning mode. This is an issue for future research to explore further. Such research might also explain how instructions on reasoning impact effort investment exactly, which cannot be inferred from our data. Process-tracing techniques such as verbal reports might be used in future studies in order to shed light on this issue.

Regarding dispositions, higher AOT scores predicted higher pre-test scores on the reasoning tasks, in line with findings by other studies (Baron, 2008; Stanovich & West, 1997, 2007). However, even though AOT scores still predicted performance on post-test scores after controlling for pre-test scores, there was no interaction with instruction conditions, indicating that instructions were equally effective for all participants. Surprisingly, NFC scores were not related to pre-test reasoning scores and negatively related to post-test reasoning scores. This finding is rather puzzling; possibly, because the effect of NFC on performance seems to be strongly related to intrinsic cognitive motivation (Cacioppo et al., 1996), the negative relation resulted from extrinsic situational pressure (i.e., the test was a compulsory part of the course). The fact that type of disposition did not interact with instruction condition suggests that *all* learners in the explicit instruction conditions benefited from the instructions and practice.

This study has some limitations. First, critical thinking was defined and measured in terms of reasoning skills with a focus on biased reasoning. It is, therefore, possible that the implicit instructions did contribute to other aspects of critical thinking that were not measured here. Secondly, the post-test immediately followed the instruction session. Therefore, it is unclear whether explicit instructions and practice would have longer-lasting effects. This is an interesting question for future research to address.

Another interesting question for future research would be what the effects would be of repeating explicit instructions and/or practice several times during a course. Finally, future research might also further investigate why prompts seem to have no additional benefits for performance, and whether such benefits might occur under different circumstances (e.g., when students would be reminded prior to the test to ask themselves similar questions as they received during practice).

Despite the above mentioned limitations, this study provided promising results for educators who wish to enhance their students' critical thinking skills in terms of unbiased

Contents

reasoning, by showing that such skills can be enhanced through simple instructional interventions –provided they are combined with practice. This is especially promising given that our experimental intervention within the existing course was of relatively short duration (i.e., a 15 min. instruction followed by approximately 20 minutes of practice of some of the reasoning tasks). Repeating instructions throughout a course or even throughout an entire economics curriculum, might have much stronger effects, and might help economics and business students to learn to avoid biased reasoning when making decisions in dynamic and complex business environments.

# Appendix A

Practiced tasks (4 out of 7 categories) * = right option

| | |
|---|---|
| Conjunction task | The coffee manufacturer has decided to extend the assortment. Below you will find some potential consequences of this decision. Which of the consequences is most likely to occur?<br><br>1) The product range will increase.*<br>2) The product range and the revenues will increase.<br>3) The product range will increase and the revenues will increase with 30%.<br>4) The product range will increase and the revenues will double.<br>5) The product range will increase and revenues will increase with 30% and the coffee manufacturer will strengthen its position in the coffee market.<br>6) The product range will increase and revenues will double, and the distribution system will improve.<br><br>Explanation[1]: Options 2, 3, 4, 5 and 6 violate the conjunction rule, as a conjunction cannot be more probable than one of its constituents. |
| Contingency task | Suppose it is decided that the assortment range of coffee flavours will be extended (i.e., new flavours will become available) in some stores but not in others. Suppose you are asked in the next year to evaluate this decision in relation to sales. The company provides you with the following information from the stores: |

| | Increased sales | No increased sales |
|---|---|---|
| Extended range of coffee flavours | 76 | 28 |
| No extended range of coffee flavours | 18 | 5 |

Table: number of stores with increased sales or not and extending coffee flavours or not.

The opinion of the company is: “Based on these data it is likely that the extension of the assortment has led to increased sales in the coffee stores”.

Do you agree with this opinion?
Option 1. fully agree
Option 2. agree
Option 3. neither agree nor disagree
Option 4. disagree*
Option 5. fully disagree*

Explanation[1]: Option 1, 2 and 3 refer to the tendency to weigh the information given in a 2 x 2 contingency table, inappropriately, by judging the importance of numerical information unequally (in this case thinking Cell A (76) is most important).

Contents

| | |
|---|---|
| Base-rate task | A large architectural firm wants to solely purchase 'sustainable coffee' from now on. The company has called an independent agency that already advised 45 large companies including governmental departments about procurement of sustainable coffee, all based on thorough research. The agency recommends purchasing coffee with a certification of the Fairtrade Labeling Organization International (FLO-I) such as coffee with the 'Max Havelaar' certificate. According to the agency this certificate offers the best guarantees that the coffee originates from honest trade. Therefore the company considers replacing Nespresso coffee, which has been used by the firm since years, for coffee with the Max Havelaar certificate. One of the branch directors notes, however, that "Sustainable coffee is ok, but Max Havelaar is less tasty coffee! Nespresso is highly appreciated and moreover Nespresso has its own quality control system for sustainability. Employees and customers love the strong, full and pure coffee with its delicious creamy layer and various flavours. Coffee is our business card. Nespresso is professional and fits to our image. Moreover, when switching to another brand all coffee machines have to be replaced and doing so is an irresponsible investment during economic decline. Also my advice is to continue with our trusted Nespresso.<br><br>What should the management of the firm best decide?<br>Option 1: Definitely follow the recommendation of the agency to purchase coffee with a certification of the (FLO-I).*<br>Option 2: Probably follow the recommendation of the agency to purchase coffee with a certification of the (FLO-I).*<br>Option 3: Probably follow the advice of the branch director to continue Nespresso.<br>Option 4: Definitely follow the advice of the branch director to continue Nespresso.<br><br>Explanation[1]: People who choose option 3 or 4 neglect the base-rate, focusing on personal or contextual "evidence" (in this case personal experiences/opinions of a branch director) in favor of more representative statistical evidence. |
| Wason selection task | From the business case you can infer that the inspectors of the coffee manufacturer will test the coffee quality. Suppose that one group of inspectors checks the origin of the coffee and attaches a label with the continent of origin on one side of the bag and another group of inspectors check the type of coffee beans and attach a label on the other side of the bag. Each of the coffee bags below has a label of the country of origin on one side and the type of coffee on the other side. The rule is: If coffee originates from Asia, then it contains Arabica coffee beans. Decide which bags you would need to turn over in order to find out whether or not the rule is being violated.<br><br>[Origin Asia] A · [Origin Africa] B · [Type Robusta] C · [Type Arabica] D<br><br>Options: AC*, AD, BC, BD<br><br>Explanation[1]: People who chose other options than AC fail to apply logical principles, tend to verify rules rather than to falsify them, or demonstrate matching bias by selecting options explicitly mentioned in the conditional statement. |

| Origin Asia | Origin Africa | Type Robusta | Type Arabica |
|---|---|---|---|
| A | B | C | D |

*Note*. [1]Participants did not receive these explanations after practicing tasks.

# 3 | Improving Critical Thinking: Effects of Dispositions and Instructions on Economics Students' Reasoning Skills[2]

2 This chapter was published as Heijltjes, A., Van Gog, T., Leppink. J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction*, *29*, 31-42, doi:10.1016/j.learninstruc.2013.07.003

This experiment investigated the impact of critical thinking dispositions and instructions on economics students' performance on reasoning skills. Participants (*N* =183) were exposed to one of four conditions: critical thinking instruction, critical thinking instruction with self-explanation prompts during subsequent practice, critical thinking instruction with activation prompts during subsequent practice, or no critical thinking instruction or prompts (control). In all conditions, practice was a within-subjects factor, some task categories present in the test were practiced on a business case, others were not. Participants in the instruction conditions significantly outperformed participants in the control condition on the immediate and delayed post-test, but only on the practiced task categories – with the exception of the self-explanations condition, which also showed a better performance than the control condition on not-practiced categories, though only on the immediate post-test. Dispositions (i.e., Actively Open-minded Thinking and Need for Cognition) predicted reasoning skills at pre-test but did not interact with instructions on post-tests performances.

Against the background of complex and rapidly changing business environments, economics students are expected to become critical thinkers (Klebba & Hamilton, 2007; Smith, 2003). Critical thinking enables students to make sound logical and unbiased decisions, and in educational situations it has been shown to lead to better learning and transfer outcomes (e.g., Facione, 2009; Halpern, 1998; Helsdingen, Van Gog, & Van Merriënboer, 2011). Therefore, it is surprising that critical thinking is rarely explicitly taught in economics curricula (Jones, 2007). This might be due to the prevailing view that immersion in business methods and strategies will lead to the spontaneous development of critical thinking skills (Jones, 2007). However, there is little evidence that such skills develop spontaneously as a consequence of instruction in a discipline (Halpern, 1999). On the contrary, research has shown that critical thinking seems to increase only if taught explicitly (e.g., Abrami et al., 2008). It is difficult, though, for educators in economics (and other domains for that matter) to derive guidelines from exisiting research regarding when, where and how to foster critical thinking throughout the curriculum (Smith, 2003). According to Abrami et al. (2008), existing studies on critical thinking instruction often lack a powerful empirical design (see also McMillan, 1987; Ten Dam & Volman, 2004; Wolcott, Baril, Cunningham, Fordham, & St.Pierre, 2002). Moreover, the definition of critical thinking or the aspects of critical thinking that are investigated tend to differ (Abrami et al., 2008).

The present study aims to empirically examine the effects of explicit critical thinking instructions on an essential aspect of critical thinking, which is highly important to decision-making in complex and high-risk situations that often occur in the field of economics: the avoidance of biased reasoning by means of more controlled rational judgment and decision-making (Evans, 2003; Smith, 2003; Stanovich & Stanovich, 2010;

Tversky & Kahneman, 1983; West, Toplak, & Stanovich, 2008). Because such biases seem inherent to human reasoning and decision-making, economists are no exception and are also prone to violating the principles of rationality (Sanfey, Loewenstein, McClure, & Cohen, 2006). This study investigates whether declarative critical thinking instruction followed by practice, either by itself or combined with self-explanation or activation prompts during practice, would enhance economics students' learning of reasoning skills. Moreover, it is investigated whether individual differences in thinking dispositions are associated with initial performance on those skills and with learning.

Before describing the research on critical thinking instructions and the rationale behind our study, we will first address the principles of critical thinking (especially in relation to biased reasoning) and its underlying cognitive mechanisms in more detail.

## Critical Thinking

The 'American Philosophical Association Delphi Panel' (Facione, 1990), characterized critical thinking for educational purposes (i.e., to support the instruction and assessment of critical thinking) as an extensive concept including both cognitive skills and dispositions. A critical thinker must, for example, be skilled at reasoning, which refers to the cognitive process of drawing conclusions from given information (Facione, 1990). Although the critical thinking literature in educational research has focused on a wide variety of topics in the past decades (see e.g., Angeli &Valanides, 2009), little attention has been paid to the avoidance of biased reasoning (West et al., 2008). Critical thinking, in terms of avoiding biased reasoning, may be classified as a subspecies of rational thinking (Facione, 2009; Stanovich, 2011; West et al., 2008). The dual processing framework unraveled the underlying cognitive processes of rational thinking (e.g., Evans, 2003, 2008, 2011). According to theories of dual processing, two distinct types of reasoning processes are at work. Type 1 processes have a rapid, automatic nature and involve little reflection. Decision-making is based on past experiences and requires little effort, which is useful and efficient in many routine situations. However, because of its automaticity, it might also result in biased thinking in other situations, unless Type 2 processes overrule these automatic responses by explicit reasoning efforts. Type 2 processes are slow, sequential in nature, and require the exclusion of attention to other matters, and therefore draw more heavily on working memory capacity.

Stanovich (2009) distinguishes Type 2 processes into reflective and algorithmic operations. The reflective mind operates at an intentional level based on dispositions such as beliefs, cognitive style, goals, and epistemic values, which affect the algorithmic mind. Research has shown that dispositions such as 'actively open-minded thinking' (AOT, i.e., the active search for evidence against one's own beliefs, plans, or goals and the

Contents

ability to weigh available evidence fairly; Baron, 2008) and 'need for cognition' (NFC, i.e., "the tendency for an individual to engage in and enjoy thinking"; Cacioppo, Petty, & Kao, 1984, p. 306) predict performance on tasks associated with rational thinking such as syllogisms, statistical reasoning, and framing (Stanovich & West, 1997, 2007; for a review on NFC, see Cacioppo, Petty, Feinstein, & Jarvis, 1996). Individuals with higher AOT scores performed better on argument evaluation (Stanovich & West, 1997) and co-variation judgment (Sá, Kelley, Ho, & Stanovich, 2005; West et al. 2008), considered more alternative possibilities (i.e., other possibilities than initially favored), asked themselves more frequently what possibility would produce the best expected outcomes, and showed less overconfidence in hasty conclusions than individuals with lower AOT scores (Baron, 2008). NFC is characterized by cognitive motivation that predicts performance on cognitive tasks. For example, individuals with higher NFC scores have been found to recall more of the information to which they are exposed, are more focused on substantive merits of the information (e.g., are more influenced by the quality of arguments of persuasive messages), generate more task-relevant thoughts which reflect the quality of arguments, make more thoughtful judgments (i.e., scrutinize and elaborate material more) and put more effort into cognitive tasks than individuals with lower NFC scores.

The algorithmic mind performs analytic and inhibitory processes that enable a person to process information in such a way that the correct actions are taken (Stanovich, 2009). Thus, the algorithmic mind has the ability to override Type 1 processes by applying knowledge of inferential rules and strategies of rational thought (e.g. probabilistic reasoning, causal reasoning and logic). Failures to override Type 1 reasoning often occur on classical heuristics and biases tasks, which tend to evoke an automatic response, while they require causal and probabilistic reasoning, assessment of covariation, a tendency to think statiscally, and to think of alternative explanations. For example Tversky and Kahneman (1983) illustrated poor probabilistic reasoning on a classical conjuntion task: "*Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of the two alternatives are more probable: 1. Linda is a bank-teller or 2. Linda is a bank teller and is active in the feminist movement.*" (p. 297). Most people are inclined to choose option 2, but indicating that option 2 is more probable than option 1 violates the conjunction rule because a conjunction cannot be more probable than one of its ingredients ($P(A\&B) \leq P(B)$).

Although heuristics and biases tasks are largely unexploited in the traditional critical thinking literature and measurement (West et al., 2008), these tasks allow for the assessment of the degree of rationality in terms of reflective and algorithmic mechanisms (Stanovich, Toplak, & West, 2008). Failures to override Type 1 processes, demonstrated on heuristics and biases tasks (Stanovich, 2009), might occur due to lack of declarative knowledge of reasoning skills or insufficient strategies to use available knowledge. Both might presumably be counteracted by instructions.

## Critical Thinking Instruction: Avoiding Biased Reasoning

Relatively few studies have focused on the questions of how to avoid biased reasoning in favor of more rational thinking (Stanovich & Stanovich, 2010), of whether rational thinking is trainable, of what type of instructions would be most helpful, and of whether instructions have a persistent impact on learners' thinking beyond the period of instruction (for reviews see Ritchart & Perkins, 2005; Stanovich, 2011).

Research that has been conducted on educational interventions revealed some evidence that explicit rule training on the law of large numbers fostered performance on base-rate problems (Fong, Krantz, & Nisbett, 1986), that explicit debiasing instructions (i.e., instructions to decouple prior beliefs and opinions from evaluation of evidence and arguments, such as prompting subjects to 'give both reasons for and reasons agianst') improved performance on argument generation tasks and syllogistic reasoning tasks (Macpherson & Stanovich, 2007). Fong et al.(1986) found that both teaching the rules and guidance to apply these rules in particular domains by examples, were both effective; however, providing examples significantly improved the use of abstract rule systems. According to Nisbett, Fong, Lehman, and Cheng (1987), on some tasks (e.g. causal and conditional tasks) reasoning improvement only had an enduring effect if abstract rule training and example training were both provided.

In the academic domains of psychology, medicine, and law, Lehman, Lempert, and Nisbett (1988) showed that teaching inferential and logical rules improved students' reasoning performances in domain-related tasks as well as on everyday life problems. In the financial domain, Larrick, Morgan, and Nisbett (1990) found that participants who had been instructed on cost-benefit principles, followed by using these principles on examples and indicating whether they agreed or disagreed with the reasoning in the examples, became more rational on reasoning (i.e., responded more normatively) compared to untrained participants and applied normative principles on a different type of problem, both immediately and after a full month, and in different contexts (i.e., transfer took place).

These findings on debiasing instructions are consistent with studies in the broader critical thinking literature (e.g., Abrami et al., 2008; Angeli & Valanides, 2009; Bangert-Drowns & Bankert, 1990), which show that general critical thinking instruction combined

Contents

with the integration of critical thinking principles into subject matter instruction (e.g., thought provoking activities) works best for improving critical thinking.

Moreover, these findings suggest that two key-factors appear to play a role in critical thinking instruction: explicit teaching of reasoning principles and room for practicing these principles. According to Bransford, Sherwood, Vye, and Rieser (1986), combining explicit instructions with practice allows students to integrate general rules into a domain specific context, which helps them to integrate, combine, and cluster information in a meaningful way. This would result in a conceptual and practical organization of knowledge that facilitates future performance on such tasks.

When both explicit instruction and practicing would be required for acquiring reasoning skills, then the question arises whether additional prompts during practice might have added value and if so, which kind of prompts? Two potential prompting methods might be useful and will be elaborated upon below: self-explanation and activation prompts.

## Self-explanation Prompts

Regarding self- explanation, Austin, Gregory, and Chiu (2008) found that prompting students to self-assess and reflect improved their performances on reasoning tasks. Participants who provided a brief written explanation of how they arrived at the rating and a rationalization of why they selected a particular answer, demonstrated a significantly better performance on critical thinking (i.e., less heuristic and more algorithmic reasoning), than participants who did not receive these prompts. Angeli and Valanides (2009) examined the impact of critical thinking teaching methods in performances on an ill-defined problem (e.g., to discuss an issue and to produce an outline for a paper) and found the highest effect size for the condition in which students discussed an issue, and reflected on their thinking combined with a short lecture about critical thinking. These findings are consistent with studies about the effectiveness of self-explanation.

Self-explanation fosters the proper use of available knowledge and skills, which engages students in active and meaningful learning while effectively monitoring their understanding (Roy & Chi, 2005). The underlying cognitive mechanisms of self-explaining have been described as "generating inferences to fill missing information, integrating information within study materials, integrating new information with prior knowledge and monitoring and repairing faulty knowledge" (Roy & Chi, 2005, p. 272). According to Lombrozo (2006), explaining is especially useful for approaches to reasoning as it promotes learning by the integration of new information with prior beliefs. Explanations help to assess the probability of for example claims in light of prior beliefs, evoke beliefs-revision, encourage learners to re-describe materials, and allow them to identify relevant principles (Willams & Lombrozo, 2010).

Moreover, self-explanation fosters generalizations. Aleven and Koedinger (2002) found that students who explained their steps during problem solving learned more effectively and generalize readily to novel situations, than students who did not explain their steps. Hence self-explanation might be a particularly effective strategy to promote both learning and transfer (e.g., Lombrozo, 2006; Renkl, 2005; Rittle-Johnson, 2006), even at a delay: Rittle-Johnson (2006) found that the effects of direct instruction combined with self-explanation facilitated learning and transfer additively and that the benefits persisted over a 2-week delay. Finally, these beneficial effects of self-explanations occur even though participants rarely receive feedback on the quality of their explanations (Matthews & Rittle-Johnson, 2009). Based on these studies it can be hypothesized that prompting self-explanation during practice of reasoning tasks would have an additional impact on acquisition of reasoning skills, as it helps not only to integrate acquired knowledge and skills (which only practising would also do), but helps to examine gaps, evoke belief-revision, and foster generalizations during practice.

### Activation Prompts

Activation prompts relate to the assumption that reasoning biases are due to a inhibition failure in working memory and not to a lack of understanding of probabilistic principles or a lack of deductive competence (Moutier, Angeard, & Houdé, 2002; Moutier & Houdé, 2003). Moutier et al. showed that inhibition training led to better performance on Wason selection tasks (Wason, 1968) and conjunction tasks (Tversky & Kahneman, 1983). For instance, they used an adaptation of the classical 'Linda task' described above, which requires inhibition of the misleading scheme (the conjunction) and activation of the relevant scheme (the probability). A control group (no training) was compared to a strictly logical training group (in which the experimenter explained the misleading scheme and admonished subjects not to fall in a trap), and an inhibition-training group. The inhibition-training group received the logical training and in addition, participants were encouraged to redirect their attention toward logically relevant information. However, even though it can be effective, inhibition training may have the unintended side effect of sensitizing the mind to other thoughts that have to be avoided (Wenzlaff & Bates, 2000). Stanovich and Stanovich (2010) also stated that suppressing the initial response (Type 1 reasoning) is only helpful when a better response is available to substitute for it. In line with findings on inhibition training, an alternative means of increasing Type 2 reasoning might be the activation of the 'weaker' cues instead of the suppression of dominant ones (i.e., evoke an attention shift from irrelevant to relevant information; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000).

Thus, based on previous studies it would be reasonable to expect that activation prompts during practice would encourage an attention shift to relevant tasks aspects, which would have an additional impact on the acquisition of reasoning skills.

In sum, an important question for education, which is addressed in the present study, is whether critical thinking in terms of unbiased reasoning can be enhanced by explicit instructions in combination with practice, and whether self-explanation or the activation prompts during practice can further enhance reasoning skills.

## The Present Study

The present study adresses the following questions: a) What is the impact of individual differences in dispositions on economics students' critical thinking as measured by their performance on judgment and reasoning tasks? b) What are the effects of instructions, practice, and prompts during practice on economics students' critical thinking as measured by their performance on reasoning tasks immediately and after a three-week delay?

Regarding dispositions, we hypothesized in line with previous research that those participants with higher scores on dispositions (i.e., AOT and NFC) would score better on an initial assessment of critical thinking skills than participants with lower scores on these dispositions (Hypothesis 1). An interesting related question is whether students with higher scores on dispositions would also benefit more from the critical instruction than students with lower scores on dispositions, or whether instruction is equally effective for all participants; this question is explored here (Question 1).

Regarding instructions, it is hypothesized that critical thinking instruction enhances performance on critical thinking skills compared to no instruction, both immediately (Hypothesis 2a) and after a three-week delay (Hypothesis 2b); however, based on prior research it can be expected that this beneficial effect would only arise when instruction is combined with practice. Therefore, practice is taken into account as a within-subjects factor, and it is hypothesized that only those tasks which will be instructed and practiced, would enhance reasoning skills (Hypothesis 2c).

If critical thinking instruction has a general effect, then performance on both practiced and not-practiced tasks should be enhanced compared to the control (i.e., no instruction) condition; however, if instruction only has a beneficial effect when combined with practice, only performance on practiced tasks should be enhanced compared to the control condition. Secondly, it is hypothesized that after instruction, combining practice with prompts (either self-explanation or activation prompts) would

Contents

foster acquisition of critical thinking skills compared to instruction and practice only (i.e., without prompts) both immediately (Hypothesis 3a) and after a three-week delay (Hypothesis 3b).

# Method

## Participants and Design

Participants were 195 part-time Economics students of a Dutch University of Applied Sciences with various educational backgrounds (in the Netherlands, Universities of Applied Sciences can be entered via various secondary education trajectories). Twelve participants had to be excluded because of missing data, leaving a final sample of 183 students (121 men, 62 women; age *M* = 29.3, *SD* = 6.5). Participants were randomly assigned to one of four conditions: 1) no critical thinking instruction (control; *n* = 40), 2) critical thinking instruction (*n* = 46), 3) critical thinking instruction plus self-explanation prompts (*n* = 47), and 4) critical thinking instruction plus activation prompts (*n* = 50). The experiment consisted of 5 phases: 1) pre-test, 2) instructions (critical thinking or unrelated topic depending on condition), 3) practice of reasoning skills on a business case, 4) immediate post-test, 5) delayed post-test. Note regarding phase 3 that in each condition, practice was a within-subjects factor (i.e., some task categories were practiced during the practice phase, others were not). In the prompting conditions, the prompts were provided during practice (phase 3).

## Materials and Procedure

**Critical thinking disposition tests.** We used a Dutch translation of the 41-item Actively Open-minded Thinking test (AOT; Stanovich & West, 2007) and the 18-item (short form) Need for Cognition questionnaire (NFC; Cacioppo et al., 1984) to measure critical thinking dispositions. Both consist of items requiring a response on a 6-point Likert scale (AOT: as in Stanovich & West, 2007; NFC: as in West et al., 2008; note that Cacioppo et al., 1984 used a 9-point scale) ranging from (1) strongly disagree to (6) strongly agree. Scores on the items are summed for AOT and for NFC separately (after reverse scoring items that are formulated negatively). Higher scores on AOT characterize a greater tendency toward open-minded thinking. The reliability of the AOT was good: the internal consistency (Cronbach's alpha) was .82. Higher scores on the NFC represent a greater tendency to engage in and enjoy effortful cognitive endeavors. Reliability of the NFC was also good: the internal consistency (Cronbach's alpha) was .80.

**Critical thinking skills tests.** The critical thinking skills tests consisted of sixteen tasks (see appendix A for an example task in each category). The tasks in the pre-test,

Contents

immediate, and delayed post-test were structurally equivalent, but surface features (cover stories) differed.

The reasoning tasks examined the heuristic tendency: 1) to be influenced by intense personal and case evidence in favor of more representative statistical evidence (two causal base-rate tasks adapted from Fong et al., 1986), 2) to base judgments on prior belief and intuition without taking sufficient account of the prior probability (two non-causal base-rate tasks adapted from De Neys & Glumicic, 2008), 3) to neglect simple and fundamental qualitative rules of probability in conjunction problems in which a conjunction cannot be more probable than one of its ingredients (two conjunction tasks, adapted from Tversky & Kahneman, 1983), 4) to shift outcomes when the same information is framed in different ways as in cases of gains and risky options (two framing tasks adapted from Tversky & Kahneman, 1981), 5) to evaluate the information given in a 2 x 2 contingency table unequally, that is, to base estimations on already experienced evidence and disregard some of the presented evidence (two covariation tasks adapted from Wasserman, Dorner, & Kao, 1990), 6) to examine the tendency to verify rules rather than to falsify them (two Wason selection tasks adapted from Stanovich, 2009, and Wason & Shapiro, 1971), and 7) to examine the tendency to evaluate the logical validity of arguments on the basis of one's prior beliefs about the truth of conclusions on syllogistic reasoning tasks (Evans, 2003; Markovits & Nantel, 1989; Sá et al., 1999). In syllogistic reasoning tasks a conclusion is drawn from two given premises or assumed propositions and participants have to indicate whether or not the conclusion follows logically from the premises. Different types of inconsistent (i.e., the validity of the conclusion is in conflict with its believability) syllogistic tasks were included (one of each type): affirming the antecedent or modus ponens (if p then q, p therefore q; valid), affirmation of the consequent (if p then q, q therefore p; invalid), denial of the antecedent (if p then q, not p therefore not q; invalid) and denial of the consequent, or modus tollens (if p then q, not q therefore not p; valid).

The content of the surface features (cover stories) of the tasks was adapted to the interests of students in the economics domain. The format of the tasks differed; a multiple-choice format with two to five answer options (depending on task characteristics) was used (the correct answer based upon rational reasoning strategies and incorrect answers related to biased reasoning). The internal consistency (Cronbach's alpha) on the three tests was .50, .70 and .73, respectively. Modest reliability for the pre-test composite score could be expected and is in line with previous research (de Bruin, Parker, & Fischoff, 2007; West et al., 2008). It should be noted that reliability on the post-tests was much higher. It is not uncommon for pre-tests to show a lower reliability as performance prior to instruction is often more random or variable than performance on post-tests after instruction.

**Critical thinking instructions.** The critical thinking instructions (see Table 1 for an overview of the study design) in the experimental conditions consisted of computer-based presentation (visual and auditory information) of 15 min. in which the features of critical thinking, its importance, the required reasoning skills, the dispositions, and the risk of biased thinking and fallacies in thinking were explained. Examples and demonstrations of all task categories were provided, referring back to the tasks seen in the pre-test, which could have allowed participants to mentally correct initially erroneous responses. As an example, the base-rate fallacy was demonstrated by a slide presenting an image of a person, not identifiable as male or female, with a guitar, beer, and some engineering tools. Next to the image the visual text appears: base-rate fallacy, group: 990 women, 10 men. At the same time the following auditory information was given: *"The base-rate fallacy is a thinking failure that occurs when the statistical distribution of a population is ignored. For example when we select a particpant randomly out of 1000 participants with 990 women and 10 men, and we will tell you that this person is called Sam, loves drinking beer and listening to hardrock music, and has graduated as mechanical engineer, then most people tend to find it most likely that this person is a man. In that case they ignore that the total group contains of 10 men only"* (NB: a different name was used in the video, a Dutch name that can also be given to both men and women, but for the sake of clarity we used an English name here). The computer controlled the sequence and pace of the video, but participants could pause, forward and rewind, although observations during the experiment suggested they hardly made use of these options The control group received a 15 min. digital video on an unrelated topic (i.e., what happens in your brain when you are in love).

**Practice phase.** After the video-instruction all participants were exposed to a business-case from an economics course containing a description of a coffee manufacturer who had to decide about marketing, quality control, extending the assortment, and the sustainability measures. Four categories of tasks (i.e., a contingency task, a conjunction task, a non-causal base-rate task, and a Wason selection task) were practiced with a similar format as the tasks used in the tests, but the cover stories of the practice tasks were derived from the business case. Participants did not receive feedback on the quality of their performance on the practice tasks.

Participants in the control condition and the critical thinking instruction only condition performed the practice tasks without additional prompts. In the instruction plus self-explanation condition, prompts were given after each task to self-explain how the answer was obtained: 'Explain by using keywords how you've come to the answer'. In the instruction plus activation prompts condition, prompts were given prior to the task to focus on the relevant factors. Participants were prompted with hints: 'search for confirmation and refutation' on the contingency task, 'think of the logical probability'

Contents

**Table 1**
*Overview of Study Design*

| Experimental conditions | Video-based critical thinking instruction | Practice 4 out of 7 categories of tasks | Additional prompts on practiced tasks |
|---|---|---|---|
| Control | No (unrelated video) | Yes | No |
| Critical Thinking Instruction | Yes | Yes | No |
| Critical Thinking + Self-explanation Prompts | Yes | Yes | Yes |
| Critical thinking + Activation Prompts | Yes | Yes | Yes |

on the conjunction task, 'think of the statistical distribution' on the non-causal base-rate task, and 'violation?' on the Wason selection task.

To measure whether mental effort invested in practicing tasks differed between conditions, a 9-point subjective rating scale ranging from (1) very, very low effort to (9) very, very high effort (Paas, 1992) was applied after each task in the practice phase. Mental effort is an indicator of actual cognitive load, and this scale is widely used in educational research (for a review see Van Gog & Paas, 2008).

## Procedure

The experiment was run in 11 sessions in a computer room at the university with 15 - 22 participants per session and all conditions represented in each session; participants had been randomly assigned to one of the conditions prior to the experiment. All of the materials were delivered in a computer-based environment that was created for this experiment and participants could work at their own pace. They first completed the pre-test and dispositions tests. Then, participants in the experimental conditions received the critical thinking instruction, while participants in the control condition watched the video on an unrelated topic. Subsequently, all participants read the business case and performed the practice tasks, with or without prompts depending on their assigned condition. Finally, they completed the immediate post-test. Three weeks later, all participants were requested via e-mail to complete the delayed post-test online.

### Data Analysis

For each correct answer on the critical thinking skills pre-test, immediate post-test, and delayed post-test, 1 point was assigned, resulting in a maximum score of 8 points for practiced tasks, and 8 points for not-practiced tasks, on each test.

Ratings on the AOT and the NFC were summed after reverse scoring negatively formulated items, resulting in a maximum score on the AOT of 246 and on the NFC of 108. Partial Eta-squared ($\eta_p^2$) values were computed to estimate the magnitude of group differences prior to treatment (i.e., pretest) in test performance and critical thinking dispositions, and Cramer's *V* was computed for group differences with regard to educational background. Further, Odds Ratios (*OR*s) were computed to estimate the magnitude of dichotomous variables that might predict dropout from immediate to delayed post-test, and Cohen's *d*-values were computed to estimate the standardized difference in average age, AOT, NFC, pre-test performance, immediate post-test performance, and invested mental effort during instruction between dropouts and non-dropouts. $R^2$ values were computed to estimate age as a predictor of post-test scores and AOT and NFC as predictors of pre-test scores. Finally, standardized regression coefficients (betas) were computed for predictor variables in the multilevel model for immediate and delayed post-test performance. To take the hierarchical structure of the data into account, not the raw test performance standard deviation (as linked to the standardized beta formula calculated in single-level multiple regression analysis) but the residual (adjusted) standard deviation of test performance was used to calculate the betas. The lowest level variance is then "the amount of variation in the outcome measure attributable to the individual observation after appropriate controls have been made" (Schagen & Elliot, 2004, p. 13). Although effect sizes tend to be slightly larger in the residual standard deviation approach as compared to a raw standard deviation approach, "such calculations are considered appropriate because they explicitly model the extent and impact of clustering in the data" (Schagen & Elliot, 2004, p. 13).

## Results

For all analyses, a significance level of .05 was used, except for post-hoc comparisons between conditions for which a significance level of .05/6 was used (note that the uncorrected *p*-values are reported). Random assignment of participants to conditions had been succesful; there were no significant diffences between conditions in critical thinking pre-test performance on practiced tasks, $F(3, 179) = 1.36$, $p = .258$, $\eta_p^2 = .02$, not-practiced tasks, $F(3, 179) = 0.44$, $p = .723$, $\eta_p^2 = .01$, critical thinking dispositions (AOT: $F(3, 179) = 0.33$, $p = .803$, $\eta_p^2 = .01$; NFC: $F(3, 179) = 1.2$, $p = .311$, $\eta_p^2 =. 02$), and educational background, $\chi^2(6) = 8.38$, $p = .212$, $V = .15$. Regarding the following analyses, it is worth noting that

Contents

the response rate on the delayed post-test was approximately 45% (*N* = 85): control (*n* = 16), critical thinking instruction (*n* = 25), critical thinking instruction plus self-explanation prompts (*n* = 21), and critical thinking instruction plus activation prompts (*n* = 23). Logistic regression to investigate whether participants who completed the immediate post-test only (i.e., dropouts) differed from participants who completed both the immediate and the delayed post-test (i.e., persisters) reveal that dropouts could not be predicted based on AOT scores ($p$ = .284, dropouts $M$ = 179.88, $SD$ = 17.09, persisters $M$ = 177.36, $SD$ 14.45, $d$ = 0.16), NFC scores ($p$ = .432, dropouts $M$ = 77.98, $SD$ = 9.11, persisters $M$ = 79.05, $SD$ = 9.44, $d$ = - 0.12), educational background ($p$ = .851, $OR$ = 1.04), gender ($p$ = .188, $OR$ = 0.65), pre-test performance ($p$ = .513, dropouts $M$ = 7.68, $SD$ = 2.31, persisters $M$ = 7.92, $SD$ = 2.55, $d$ = - 0.01), immediate-post-test performance ($p$ = .283, dropouts $M$ = 9.11, $SD$ = 3.31, persisters $M$ = 9.60, $SD$ = 2.76, $d$ = - 0.16), instruction conditions ($p$ = .862, $OR$ = 0.95) or invested mental effort during instructions ($p$ = .342, dropouts $M$ = 3.65, $SD$ = 1.30, persisters $M$ = 3.83, $SD$ = 1.20, $d$ = - 0.14). Only age was a significant predictor of dropout ($p$ = .004, dropouts $M$ = 30.6, $SD$ = 7.35, persisters $M$ = 27.9, $SD$ = 5.12, $d$ = 0.43), indicating a small to medium effect (Cohen, 1992); however this effect seems of minor importance as age was not a predictor of immediate post-test scores ($p$ = .089, $R^2$ = .02) or delayed post-test scores ($p$ = .076, $R^2$ = .04). All in all, the $p$-values in combination with the measures of effect size indicate that the groups who did and did not complete the delayed posttest were comparable on the reported variables prior to treatment.

To explore whether invested mental effort during practice differed among instruction conditions an ANOVA was conducted, which showed no significant diffences among conditions, $F(3, 179) = 1.30$, $p = .276$, $\eta_p^2 = .02$.

## Dispositions and Pre-test Performance

To test the hypothesis that participants with higher scores on dispositions (i.e., AOT and NFC) would score better on the pre-test of critical thinking skills than participants with lower scores on these dispositions (Hypothesis 1), a multiple regression analysis was performed. On the dispositions tests, data from 11 participants were lost due to a technical error. Their scores were replaced with the average sample scores. In line with our hypothesis, the dispositions AOT and NFC significantly predicted pre-test critical thinking skills, $F(2, 180) = 11.55$, $p < .001$, $R^2 = .11$, indicating a medium effect (Cohen, 1992). Regarding the standardized coefficients the AOT scores showed a slightly higher impact, $\beta = .22$, $t(180) = 2.95$, $p = .004$, than the NFC scores, $\beta = .20$, $t(180) = 2.64$, $p$ = .009, indicating a small to medium effect (Lipsey & Wilson, 2001).

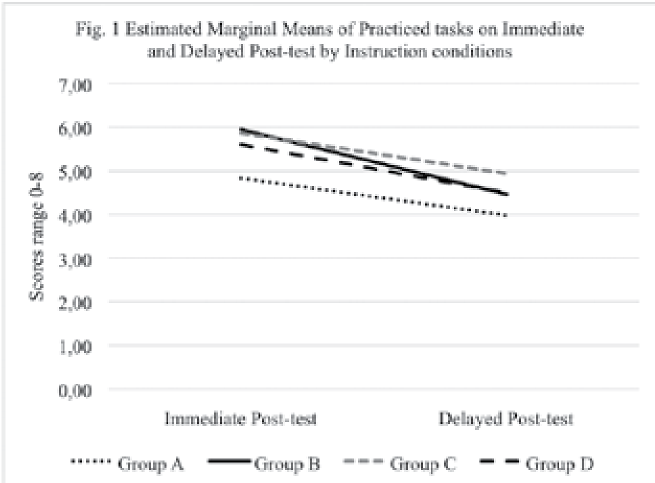
## Effects of Instructions and Dispositions on Immediate and Delayed Post-test Performance

To include not only participants who completed the immediate *and* the delayed test but to use as much information as possible, a multilevel regression model was used. The following fixed factors were included in the model consistently: pre-test scores (as covariate), dispositions scores (i.e., AOT and NFC), invested mental effort scores during instruction, instruction conditions, practice (i.e., practiced tasks versus not practiced tasks), and test moment (i.e., immediate post-test scores and delayed post-test scores). A student-level random intercept and random slope for practice were included in the model. No abnormal departures from normality or outliers were found. Figure 1 and 2 display the mean reasoning scores of practiced and not-practiced tasks in each condition on the immediate and delayed post-test. Table 5 presents the adjusted means along with their standard errors and 95% confidence intervals for every combination of condition by practice by test moment, keeping all quantitative covariates constant at their mean value.

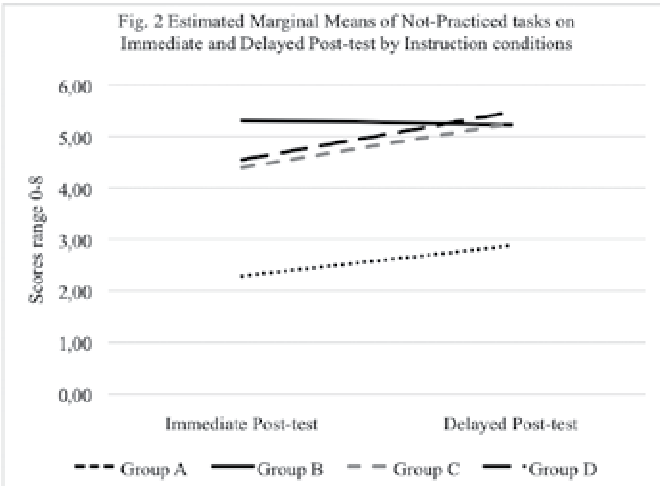
Regarding the explorative question of whether students with higher scores on dispositions would also benefit more from the critical instruction than students with lower scores on dispositions, or whether instruction would be equally effective for all participants (Question 1), Table 2 shows that the dispositions and critical thinking performance were correlated significantly even after instruction (i.e., mainly between AOT and post-test scores), however the multilevel regression analysis, using Satterthwaite approximation for the degrees of freedom in SPSS version 21 (Table 3), revealed that no interactions were found between scores on dispositions (i.e., AOT and NFC) and instructions, between dispositions and practice, or between dispositions and test moment.

The hypotheses that instruction would have an effect on post-test performance (Hypothesis 2a) when combined with practice (Hypothesis 2c), that prompts during practice would further enhance the effects of instructions and practice (Hypothesis 3a), and that these effects would remain the same after a three-week delay (Hypothesis 2b and 3b), were also tested by means of the multilevel regression analysis (Table 3). This analysis showed significant main effects of instruction condition and test moment, and significant interaction effects between instruction condition and practice, and between test moment and practice. Table 4 shows the *B*-values for the main and interaction effects and the effect sizes (betas). Note that the intercept is based on not-practiced tasks of the immediate post-test in the control condition (i.e., group A). The betas for pre-test reasoning scores and invested mental effort indicate that pre-test reasoning score has a strong positive effect on immediate and delayed post-test performance, while invested

Contents




*Note.* Group A: control, Group B: critical thinking instruction, group C: critical thinking instruction + self-explanation prompts, Group D: critical thinking + activation prompts. Covariates in the model kept constant on their mean value: Scores on pre-test reasoning tasks, invested mental effort during instruction, Actively Open-minded Thinking, and Need for Cognition.




*Note.* Group A: control, Group B: critical thinking instruction, group C: critical thinking instruction + self-explanation prompts, Group D: critical thinking + activation prompts. Covariates in the model kept constant on their mean value: Scores on pre-test reasoning tasks, invested mental effort during instruction, Actively Open-minded Thinking and Need for Cognition.

mental effort during instruction appears to have a small to moderate negative effect on performance. In line with the significance tests presented in Table 3, the significance tests and betas in Table 4 appear to indicate that NFC and AOT hardly influence post-test performance. Further, the condition by practice interaction appears to be a strong

**Table 2**
*Correlations between Critical Thinking Performance and Dispositions (AOT and NFC) on Pre-test, Immediate Post-test and Delayed Post-test of Practiced and Not-practiced tasks.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. AOT | - | | | | | | |
| 2. NFC | .321*** | - | | | | | |
| 3. Pre-test Practiced tasks | .220** | .173* | - | | | | |
| 4. Pre-test Not-practiced tasks | .244** | .263*** | .364*** | - | | | |
| 5. Immediate Post-test Practiced-tasks | .323*** | .192** | .379*** | .317*** | - | | |
| 6. Immediate Post-test Not-practiced tasks | .259*** | .114 | .341*** | .502*** | .440*** | - | |
| 7. Delayed Post-test Practiced tasks | .330** | .123 | .240* | .198 | .653*** | .434*** | - |
| 8. Delayed Post-test Not-practiced tasks | .222* | .106 | .283** | .430*** | .355** | .655*** | .428*** |

*Note*. AOT: Actively Open-minded Thinking scores, NFC: Need for Cognition scores, *** Significant at the 0.001 level, **Significant at the .001 level, *Significant at the 0.05 level.

effect; in the control condition, immediate post-test performance is considerably lower on practiced tasks (notice the negative beta), while in the other three conditions immediate post-test performance is much higher on the practiced tasks than on the not practiced tasks.

To investigate the interaction effect between condition and practice further, simple effect analyses taking into account the factors practice (i.e., practiced versus not-practiced tasks) and test moment (i.e., immediate and delayed post-test), revealed that participants in the instruction conditions outperformed the controls on practiced tasks on the immediate post-test (all *ps* $< .001$) and on the delayed post-test (all *ps* $<$ .002). LSD post-hoc tests on practiced tasks on the immediate and delayed post-test revealed no significant differences between instruction conditions (immediate test: all *ps* $> .484$, delayed test: all *ps* $> .484$). On not-practiced tasks, only participants in the instruction condition with self-explanation prompts outperformed those in the control condition on the immediate post-test ($p < .001$) but not on the delayed post-test ($p =$ .199). LSD post-hoc tests showed that other comparisons between conditions on not-practiced tasks were not significant on either the immediate post-test (all *ps* $> .020$) or

Contents

**Table 3**
*Fixed Effects for Predictors of Reasoning Scores.*

| Parameter | *df* | *F* | *p* |
|---|---|---|---|
| Intercept | 1,186.06 | 43.63 | < .001 |
| Conditions | 3,289.58 | 3.94 | .009 |
| Practice | 1,302.88 | 3.17 | .076 |
| Test moment | 1,255.56 | 51.95 | < .001 |
| Conditions x Practice | 3,302.20 | 10.85 | < .001 |
| Conditions x Test moment | 3,254.98 | 1.10 | .351 |
| Conditions x Practice x Test moment | 3,259.66 | .28 | .842 |
| Practice x Test moment | 1,260.12 | 60.93 | < .001 |
| Pre-test reasoning scores | 1,166.67 | 82.39 | < .001 |
| Invested mental effort | 1,173.95 | 4.62 | .033 |
| NFC | 1,262.90 | .11 | .747 |
| AOT | 1,267.42 | 1.96 | .163 |
| NFC x Test moment | 1,255.12 | 1.04 | .309 |
| AOT x Test moment | 1,270.45 | .85 | .358 |
| NFC x Conditions | 3,166.92 | .96 | .414 |
| AOT x Conditions | 3,178.35 | 1.95 | .124 |
| NFC x Practice | 1,217.96 | 1.50 | .223 |
| AOT x Practice | 1,235.33 | 2.55 | .112 |

*Note.* Intercept based on not-practiced tasks on the immediate post-test; Practice: practiced tasks (code '1') versus not-practiced tasks (code '0'); Test moment: immediate post-test (code '0') versus delayed post-test (code '1'). AOT: Actively Open-minded Thinking, NFC: Need for Cognition, Conditions: control, critical thinking instruction, critical thinking instruction + self-explanation prompts, critical thinking + activation prompts.

the delayed post-test (all *ps* > .030). Note that the uncorrected *p*-values are reported, but these are no longer significant after correction for multiple tests (i.e., .05/6).

Finally, the practice by test moment interaction (Table 4) indicates that there is a decrease in performance on the practiced tasks from immediate to delayed post-test whereas there is no change in performance on not-practiced tasks; the beta of practice by test moment suggests that this is a strong effect. Note though (Table 3), that there was no significant interaction with condition.

Contents

**Table 4**
*Fixed Effects Estimates and Covariance Estimates of Reasoning Scores.*

| Parameter | *B (SE)* | *df* | *t* | *p* | *Effect size (beta)* |
|---|---|---|---|---|---|
| Level 1 | Fixed effects | | | | |
| Intercept | 1.98 (0.43) | 208.38 | 4.62 | < .001 | |
| Group B | .52 (0.29) | 289.85 | 1.81 | .072 | 0.48 |
| Group C | .99 (0.29) | 289.48 | 3.44 | < .001 | 0.91 |
| Group D | .55 (0.28) | 288.78 | 1.94 | .054 | 0.51 |
| Practice | -1.09 (0.29) | 302.01 | -3.74 | < .001 | -1.01 |
| Test moment | .92 (0.35) | 262.03 | 2.66 | .008 | 0.85 |
| Practice x Test moment | -1.49 (0.48) | 267.24 | -3.11 | .002 | -1.38 |
| Group B x Practice | 1.86 (0.40) | 302.10 | 4.66 | < .001 | 1.72 |
| Group C x Practice | 1.41 (0.40) | 301.15 | 3.53 | < .001 | 1.30 |
| Group D x Practice | 2.07 (0.39) | 302.64 | 5.31 | < .001 | 1.91 |
| Group B x Test moment | .63 (0.45) | 256.05 | 1.40 | .161 | 0.58 |
| Group C x Test moment | -.02 (0.46) | 259.51 | -.05 | .962 | -0.02 |
| Group D x Test moment | -.11 (0.45) | 258.31 | -.24 | .807 | 0.10 |
| Group B x Practice x Test moment | .01 (0.62) | 259.53 | .02 | .983 | 0.01 |
| Group C x Practice x Test moment | -.28 (0.64) | 264.55 | -.44 | .664 | -0.26 |
| Group D x Practice x Test moment | -.44 (0.63) | 264.58 | -.70 | .487 | -0.40 |
| Pre-test reasoning scores | .32 (0.04) | 166.67 | 9.08 | < .001 | 0.71 |
| Invested mental effort | -.14 (0.07) | 173.95 | -2.15 | .033 | -0.16 |
| NFC | .01 (0.02) | 198.80 | .52 | .602 | 0.09 |
| AOT | -.01 (0.01) | 213.56 | -1.18 | .241 | -0.18 |
| Group B x AOT | .03 (0.02) | 169.30 | 1.78 | .076 | 0.40 |
| Group C x AOT | .04 (0.01) | 186.69 | 2.07 | .040 | 0.45 |
| Group D x AOT | .03 (0.01) | 177.66 | 1.91 | .058 | 0.40 |
| Group B x NFC | -.02 (0.03) | 163.57 | -.60 | .552 | -0.14 |
| Group C x NFC | -.04 (0.03) | 175.93 | -1.46 | .146 | -0.33 |
| Group D x NFC | -.00 (0.03) | 167.16 | -.09 | .932 | -0.02 |
| AOT x Practice | .01 (0.01) | 235.33 | 1.60 | .112 | 0.19 |
| NFC x Practice | .02 (0.01) | 217.96 | 1.22 | .223 | 0.15 |
| AOT x Test moment | .01 (0.01) | 270.45 | .92 | .358 | 0.11 |
| NFC x Test moment | -.01 (0.01) | 255.12 | -1.02 | .309 | -0.11 |
| Level 2 | Random parameters | | | | |
| | *cov (SE)* | *Wald Z* | | *p* | |
| Intercepts participants | .53 (0.13) | 4.25 | | < .001 | |
| Practice participants | 1.05 (0.26) | 4.02 | | < .001 | |
| Residuals | 1.17 (0.11) | 10.21 | | < .001 | |

*Note.* Intercept based on not-practiced tasks on the immediate post-test; Practice: practiced tasks (code '1') versus not-practiced tasks (code '0'); Test moment: immediate post-test (code '0') versus delayed post-test (code '1'). Conditions: group A: control condition, group B: CT instruction, group C: CT instruction + self-explanation prompts, group D: CT-Instruction + activation prompts; AOT: Actively Open-minded Thinking.
Effect sizes: 0.1 small; 0.25 medium, 0.40 large (Lipsey &Wilson, 2001).

Contents

**Table 5**
*Estimated Marginal Means, Standard Errors and 95% Confidence Intervals of Practiced and Not-practiced Task Categories on the Immediate and Delayed Post-test by Groups.*

| | | Test-moment | | | | | |
|---|---|---|---|---|---|---|---|
| | | Immediate Post-test | | | Delayed Post-test | | |
| | Group | *M* | *SE* | *CI* | *M* | *SE* | *CI* |
| Practiced tasks | A | 4.83 | .34 | [4.17, 5.50] | 3.98 | .23 | [3.53, 4.43] |
| | B | 5.95 | .27 | [5.42, 6.49] | 4.64 | .21 | [4.04, 4.88] |
| | C | 5.87 | .29 | [5.29, 6.44] | 4.94 | .21 | [4.52, 5.36] |
| | D | 5.60 | .28 | [5.05, 6.16] | 4.50 | .21 | [4.09, 4.90] |
| Not-practiced tasks | A | 2.29 | .34 | [1.63, 2.96] | 2.88 | .23 | [2.43, 3.33] |
| | B | 5.31 | .27 | [4.77, 5.84] | 5.23 | .21 | [4.81, 5.65] |
| | C | 4.40 | .29 | [3.82, 4.98] | 5.26 | .21 | [4.83, 5.68] |
| | D | 4.55 | .28 | [3.40, 5.11] | 5.47 | .21 | [5.07, 5.87] |

*Note.* Group A: control, Group B: instruction only, Group C: instruction and self-explanation prompts, Group D: instruction and activation prompts. Covariates in the model kept constant on their mean value: Scores on pre-test reasoning tasks, invested mental effort during instruction, Actively Open-minded Thinking, and Need for Cognition.

## Discussion

The aim of this study was to examine a) the impact of individual differences in dispositions and b) effects of instructions, practice, and prompts on economics students' critical thinking as measured by their performance on reasoning tasks on an immediate and delayed post-test.

In line with our first hypothesis (Hypothesis 1), the results clearly indicate that those participants with higher scores on dispositions (i.e., AOT and NFC) scored significantly better on the initial assessment of critical thinking skills than participants with lower scores on dispositions. These results converge with findings in other studies on argument evaluation (Stanovich & West, 1997), syllogistic reasoning (Macpherson & Stanovich, 2007), and covariation judgment (Sá et al., 2005; West et al., 2008). Regarding the question of whether students with higher scores on dispositions would benefit more from the critical thinking instruction than students with lower scores on dispositions (Question 1), our results showed no significant interaction effects between instruction conditions and dispositions (i.e., AOT and NFC). Thus, it seems that despite the relationship between dispositions and pre-test performance, all students benefitted equally from instructions in terms of post-test performance; students who score higher on dispositions did not benefit more from instructions than students who score lower on dispositions or vice versa. These results indicate that the active search for evidence against one's own beliefs, plans, or goals, and the ability to weigh available evidence fairly (i.e., AOT; Baron, 2008), and the intrinsic cognitive motivation (i.e., NFC; Cacioppo et al., 1996), which are both important features of the reflective mind (Stanovich, 2011), appear to regulate reasoning, but not learning to reason. These explorative findings suggest that dispositions might be overruled by extrinsic factors in the educational context; declarative instruction to search for general and underlying principles and the instruction to give considerable thoughts to the instructions seems to have cancelled out any influence of AOT and NFC on learning. These results indicate that instructions were effective for all participants regardless of their disposition scores.

Our second hypothesis was that critical thinking instruction would enhance performance on critical thinking skills compared to no instruction, both immediately (Hypothesis 2a) and after a three-week delay (Hypothesis 2b), though only when combined with practice (Hypothesis 2c). Indeed, it seems that the combination of instructions with practice is crucial and has a large effect: on both the immediate and the delayed post-test the improvements on reasoning skills compared to the control condition were found on task categories that were practiced with the business case, but not on task categories that had not been practiced. There was one exception: on the immediate test, the self-explanation condition performed better than the control condition on not-practiced tasks. We will return to this finding below, when discussing effects of prompting (Hypothesis 3). Our results are in line with and extend findings from previous studies in which it was found that explicit instructions improved critical thinking (e.g., Abrami et al., 2008; Angeli & Valanides, 2009), that rational thinking to override Type 1 processes is trainable through explicit instruction (e.g., Nisbett et al., 1987) and that even short instructions can have a persistent impact on becoming more rational at reasoning (Larrick et al., 1990). On the other hand, our results also show that critical thinking instruction without practice was not sufficient to lead to sustained effects and that practice had an added value for reasoning instruction. Practice seems to play a key role and presumably leads to better learning by allowing participants to integrate and elaborate information from the instructions (e.g., Bransford et al., 1986) with tasks in the economics context. Interestingly, mental effort during practice did not differ between conditions, and combined with the differences in post-test scores on practiced tasks, this suggests that the processes engaged in during practice differ as a function of instructions, but that the cognitive demands imposed by those processes do not differ (i.e., practice becomes more efficient).

The results only lend partial support to our third hypothesis that after instruction, combining practice with prompts, would foster acquisition of critical thinking skills compared to instruction and practice only both immediately (Hypothesis 3a) and after

Contents

a three-week delay (Hypothesis 3b). Activation prompts, which we expected to be able to affect critical thinking performance by redirecting attention to relevant cues, did not lead to better test performance than instructions only. This finding might indicate that instructions were sufficient to redirect attention and inhibit initial automatic responses during practice. Self-explanation prompts, which we expected to foster the proper use of available knowledge (Roy & Chi, 2005) also did not foster performance on practiced reasoning tasks compared to instructions only. Interestingly, however, participants in this condition performed better on not-practiced reasoning tasks on the immediate post-test than the control condition, suggesting that prompting self-explanations established a kind of transfer from practiced to not-practiced tasks, at least in the short run (this effect was no longer present on the delayed test). This result converges with studies on the benefits of self-explanation of problem solving tasks that showed that self-explanation can foster transfer (e.g., Aleven & Koedinger, 2002; Lombrozo, 2006; Renkl, 2005; Rittle-Johnson, 2006). However, given that the difference with the other instruction conditions was not significant, and that the effect was no longer present on the delayed test, this effect should be interpreted with caution and should be replicated in future research. It might also be worthwhile to attempt to deepen instructional effects of self-explanation in future studies by explicitly teaching it as a meta-strategy or by providing feedback on self-explanations, as this might further enhance the effects of self-explanation on critical thinking and potentially on transfer to not-practiced tasks.

In sum, based on these findings, we can conclude that dispositions only have an impact on pre-test reasoning performance but not on learning of reasoning skills. Secondly we can conclude that explicit reasoning instructions should be combined with opportunities for practice to have an effect on critical thinking performance. Thirdly, prompting self-explanation during practice might be an interesting instructional method as we found some indications of transfer in terms of performance benefits on not-practiced reasoning tasks compared to the control condition; however, further research would be necessary to establish the merits of this method for teaching reasoning skills.

This study has some limitations. First, for practical reasons it was not possible to administer the delayed post-test at the institute as the rest of the experiment, however the web-based environment that was used for the delayed post-test was the same as the one used during the session at the institute. Secondly, despite the fact that students could complete the delayed test wherever they wanted, there was a rather high attrition rate on the delayed test. However, it should be noted that participants who did complete the delayed test did not differ from participants who did not on pre-test performance, dispositions, educational background, gender, invested mental effort during instruction, and performance on the immediate test. Third, we did not include conditions that received the prompts during practice but no critical thinking instruction. The reason for that was that it can be questioned whether prompts would have a beneficial effect for students lacking knowledge about the tasks (i.e., without instructions first). As Stanovich and Stanovich (2010) stated, on the tasks we used, suppressing the initial response (Type 1 reasoning) is only helpful when a better response is available to substitute for it. And regarding self-explanations it has been shown that without any knowledge, it would be unlikely that students would be able to provide deep, principled explanations, which have been shown to be the most effective (Renkl, 1997). Under conditions of very low prior knowledge, instructional explanations have been shown to be most effective (Renkl, 2002). The instructions provided students with a basis for responding to the prompts, which they would not have had otherwise. Given the low effectiveness of prompting even after instructions, it seems highly unlikely that prompting without instructions first would have had an effect. Nevertheless, we did not directly test this, and future research might resolve the question of whether self-explanation or activation prompts have an impact on reasoning improvements without prior critical thinking instruction. Finally, it should be kept in mind that this study defined critical thinking in terms of reasoning skills, and that the findings therefore may not generalize to other definitions of critical thinking.

Despite these limitations, this study provided promising results for (economics) educators who wish to enhance their students' critical thinking skills, by showing that such skills can be enhanced through relatively simple explicit instructions –provided they are combined with practice. The findings are especially promising given that this experimental study was of relatively short duration; teaching critical thinking throughout the economics curriculum might help students to learn to avoid biased reasoning and better prepare them for decision-making in dynamic and complex business environments.

Contents

# Appendix

Example of each category of tasks used on critical thinking tests (* = right option; 1 = category practiced; 2 = category not practiced)

| | |
|---|---|
| Conjunction task (1) | A kitchen manufacturer wants to reposition itself in the kitchen industry. The goal is to increase the market share to 10%. Based on a strength/weakness analysis, measures have been taken to improve market share, pay more attention to the entrepreneurship of the employees, and conduct a more value-oriented communication campaign towards customers.<br>Which option is most likely?<br>Option 1. The market share of the kitchen manufacturer will increase by 3%.*<br>Option 2: The market share of the kitchen manufacturer will increase by 5%.<br>Option 3: The market share of the kitchen manufacturer will increase by 3%, and the satisfaction of customers and employees will improve.<br>Option 4: The market share of the kitchen manufacturer will increase by 5%, and the satisfaction of customers and employees will improve.<br>Explanation: Options 2, 3 and 4 violate the conjunction rule as a conjunction cannot be more probable than one of its constituents. |
| Contingency task (1) | An insurance company claimed that too much time was spent on advising customers without any results. Costs and benefits were not balanced. Applying a new advising system should improve this situation. The new system was used on a part of the customers. A student was assigned to evaluate the effect of the new system to determine the systems' time efficiency. The student obtained the information below and concludes that the application of the new system resulted in saving of time. |

| | New system used | New system not used |
|---|---|---|
| Saving time | 325 | 260 |
| No time saving | 90 | 55 |

Table: number of customers on which the system was applied or not and time was saved or not.
Do you agree with the student?
Option 1. fully agree
Option 2. agree
Option 3. neither agree nor disagree
Option 4. disagree*
Option 5. fully disagree*

Explanation: Option 1, 2 and 3 refer to the tendency to evaluate the information given in a 2 x 2 contingency table unequally.

Contents

| | |
|---|---|
| Causal base-rate task (1) | A renowned regional company has two vacancies for junior economists. This company has very good experiences with economics master graduates from a specific University for over 15 years. The company management will continue this policy of hiring graduates from this University. A new personnel manager, however, suggests attracting economics graduates of a Polytechnic University. The manager argued that he gained outstanding experiences in a business where he worked formerly with a polytechnic graduate who functioned at an excellent level from the outset. The personnel manager believes that polytechnic graduates work equally well as master graduates from the specific University.<br>What should the management of the company best decide?<br>Option 1: Definitely chose for master graduates of the specific University.*<br>Option 2: Probably chose for master graduates of the specific University.*<br>Option 3: Probably chose for graduates of the Polytechnic University.<br>Option 4: Definitely chose for graduates of the Polytechnic University.<br><br>Explanation: People who choose option 3 or 4 neglect the base-rate, for example motivated by personal and case evidence in favor of more representative statistical evidence. |
| Non - causal base-rate task (2) | A study had 1000 participants. Among the participants there were 25 men and 975 women. Sam is a randomly chosen participant in this study. Sam is 23 years old, graduated as mechanical engineer and enjoys going out with friends, likes drinking beer and is a fan of hard rock music.<br>Which option is most probable?<br>Option 1: Sam is a man.<br>Option 2: Sam is a woman.*<br><br>Explanation: Option 1 demonstrates the tendency to base judgments on prior belief and intuition rather than on logical reasoning (i.e., taking into account the prior probability). |
| Framing task (2) | Imagine that you face the following pair of concurrent decisions: First examine both decisions, then indicate the options you prefer.<br>A sure gain of 480 euro.<br>25% chance to gain 2000 euro and 75% chance to gain nothing.<br><br>A sure loss of 1500 euro.<br>75% chance to lose 2000 and 25% chance to lose nothing.<br><br>Options: AC, BC*, AD, BD<br><br>Explanation: The paired choices were presented together but the problem was 'framed' as a pair of separate choices. The combination B&C is superior: 25% chance to win 500 and 75% chance to lose 1500. For example, with A&D there is 25% chance to win 480 and 75% chance to lose 1520. In case of gains people tend to chose risk aversive (option A) and in case of losses to chose risk seeking (option D). |

Contents

| | |
|---|---|
| Wason selection task (1) | Each of the tickets below has a destination on one side and an airline on the other side. Here is a rule: If 'Barcelona' is on one side of the ticket, then 'Ryanair' is on the other side of the ticket. Your task is to decide which tickets you would need to turn over in order to find out whether or not the rule is being violated. |

| Destination Barcelona | Destination Madrid | Airline Ryanair | Airline Transavia |
|---|---|---|---|
| A | B | C | D |

Options: AC, AD*, BC, BD

Explanation: People who chose other options than AD probably fail to apply logical principles, verify rules rather than to falsify them, or demonstrate matching bias by selecting options explicitly mentioned in the conditional statement.

**Syllogistic reasoning tasks (2)**

The category syllogistic reasoning tasks exist of 4 types of tasks.
In the following assignments, you will be given two premises, which you must assume are true. A conclusion from the premises then follows. You must decide whether the conclusion follows logically from the premises or not.

Premises:
All oil companies are quoted companies.
Shell is a quoted company.
Conclusion: Shell is an oil company.
Option 1. Conclusion follows logically from premises.
Option 2. Conclusion does not follow logically from premises.*
Explanation:
Syllogism type: Affirmation of consequent, invalid

Premises:
All mammals walk.
Dolphins are mammals.
Conclusion: Dolphins walk.
Option 1. Conclusion follows logically from premises.*
Option 2. Conclusion does not follow logically from premises.
Explanation:
Syllogism type: Affirmation of antecedent/Modus Ponens, valid

Premises:
All things that move love water.
Cats do not love water.
Conclusion: Cats do not move.
Option 1. Conclusion follows logically from premises.*
Option 2. Conclusion does not follow logically from premises.
Explanation:
Syllogism Type: Denial of consequent/ Modus Tollens, valid

Premises:
All oil countries are rich.
Belgium is not an oil country.
Conclusion: Belgium is not rich.
Option 1. Conclusion follows logically from premises.
Option 2. Conclusion does not follow logically from premises.*
Explanation:
Syllogism type: Denial of antecedent, invalid

# 4 Unravelling the Effects of Critical Thinking Instructions, Practice, and Self-explanation on Students' Reasoning Performance[3]

3 This chapter is submitted for publication as Heijltjes A., Van Gog, T., Leppink. J., & Paas, F. (2013). *Unraveling the effects of critical thinking instructions, practice, and self-explanation on students' reasoning performance*. Manuscript submitted for publication.

Acquisition of critical thinking (CT) skills is considered an important goal in education, but it is still unclear which specific instructional techniques are effective for fostering it. The main aim of this study was to unravel the impact of CT instructions, practice, and self-explanation prompts during practice, on students' reasoning skills that are prone to bias. Another aim was to replicate findings regarding the influence of dispositions on reasoning skills prior to and after instructions, and to explore the relationship between reasoning performance, confidence, and invested mental effort prior to and after instructions. Participants were 152 first year economics students of a Dutch University of Applied Sciences. In a pretest posttest control group design, participants were exposed to one of six conditions: A) CT-instruction text, B) CT-instruction text combined with CT-practice, C) CT-instruction text combined with CT-practice and self-explanation prompts, D) unrelated text, E) unrelated text and CT-practice, F) unrelated text and CT-practice and self-explanation prompts. Only participants exposed to CT-instruction (conditions A, B, and C) improved their reasoning skills from pre-test to post-test; practice and self-explanation prompts did not improve reasoning compared to instructions only. Dispositions (i.e., Actively Open-minded Thinking) correlated positively with pre- and post-test reasoning scores; however, the instructions were equally effective for all participants. Confidence scores correlated negatively with invested mental effort. Instructions affected invested mental effort but not confidence ratings on the post-test.

Despite the fact that the acquisition of critical thinking (CT) skills has been an educational ideal for more than a century (Dewey, 1997), it is still unclear how these skills can best be taught (Davies, 2013; Halpern, 1999). An essential aspect of CT is the ability to engage in unbiased reasoning by means of controlled rational judgment and decision-making (Facione, 1990; West, Toplak, & Stanovich, 2008). Teaching unbiased reasoning is essential for dynamical and complex professional environments such as economics (Smith, 2003), legal judgment (Rachlinski, 2004), and medicine (Mamede, et al. 2010), in which biased reasoning can lead to erroneous decisions with severe financial, emotional, or even lethal consequences (e.g., Koehler, Brenner, & Griffin, 2002). In contrast to the common believe among teachers, students' critical thinking skills do not develop spontaneously in students as a side-effect of higher education (Davies, 2013; Jones, 2007): research has shown that many undergraduates do not seem to improve in critical thinking during their college years (Arum & Roksa, 2011) and graduates from higher education frequently show poor thinking strategies (Flores, Matkin, Burbach, Quinn, & Harding, 2012).

However, like for CT-skills in general (e.g., Niu, Behar-Horenstein, & Garvan, 2013), it is not entirely clear how the ability to engage in unbiased reasoning, can best be taught (Larrick, 2004; Ritchart & Perkins, 2005). The present study aims to contribute to our knowledge about teaching CT, by disentangling the effects of instructions, practice, and self-explanations on students' performance on reasoning tasks.

## Research on Training Reasoning: Explicit Instructions, Practice, and Self-explanations

Explicit CT instructions have been shown to have a beneficial effect on CT skills in general (Abrami et al., 2008; Bangert-Drowns & Bankert, 1990; Marin & Halpern, 2011), and also specifically on reasoning skills (Larrick, Morgan, & Nisbett, 1990; Macpherson & Stanovich, 2007; Nisbett, Fong, Lehman, & Cheng, 1987). Those instructions direct learners' attention to the central concepts and principles of the learning task (Lee & Anderson, 2012). Moreover, it seems that the beneficial effect of explicit instruction on thinking skills can be increased by combining it with practice on a domain-specific task (Perkins & Salomon, 1989). Practicing reasoning tasks in a familiair context may allow students to integrate knowledge from the instructions with their prior knowledge and to engage in deeper processing (Bransford, Sherwood, Vye, & Rieser, 1986). The positive effect of combining explicit instructions and practice on economics' students reasoning skills was recently confirmed in a study by Heijltjes, Van Gog, Leppink, and Paas (2014). They found that neither instructions without practice, nor practice without instructions was sufficient to lead to sustained effects. However, in that study participants in the instruction plus practice condition probably spent more time on the learning phase, and the question can therefore be raised whether or not additional, meaningful time-on-task was responsible for the beneficial effects on reasoning skills.

Moreover, Heijltjes et al. (2014) investigated whether prompting students to self-explain during practice, would improve reasoning performance. Prompting self-explanations after or during explicit instruction was shown to foster the proper use of available knowledge (Roy & Chi, 2005), problem solving transfer (e.g., Lombrozo, 2006), and beliefs-revision (Willams & Lombrozo, 2011). Heijltjes et al. found that adding self-explanation prompts during practice did not enhance immediate post-test performance on practiced tasks, but did enhance performance on not-practiced tasks, suggesting that self-explaining aided transfer –at least short term, as this effect disappeared on the delayed post-test. Note that in the self-explanation condition, tasks that were not practiced had still been instructed. Therefore, if students attempted to self-explain also on the post-test, they had knowledge from the instructions that allowed them to engage effectively in explaining. Consequently, the question is whether self-explaining would also aid transfer without any instructions about the tasks, which leaves nothing but prior knowledge to base explanations upon.

Contents

Therefore, the present study builds upon the findings by Heijltjes et al. (2014) and primarily aims to unravel the impact of CT instructions, practice, and self-explanations on economics' students reasoning.

## The Present Study

An overview of the study design with the conditions A to F is provided in Table 1. It is hypothesized that explicit instructions would be necessary to enhance performance on reasoning tasks (Hypothesis 1a: condition ABC > DEF) and that practice only, either with or without self-explanation prompts, will not lead to enhanced reasoning performance compared to controls (Hypothesis 1b: condition D = EF). These hypotheses apply to tasks that were instructed and/or practiced; the study by Heijltjes et al. (2014) found no transfer of the effects of instruction and practice to other types of tasks.

In the study by Heijltjes et al. (2014), only a combination of explicit instruction with practice improved economics students' reasoning performance. However, since practice can also result in additional time-on-task, it is still an open question whether an alternative approach that increases time on task and mindful processing would be as effective. Condition A implemented such an approach: restudying the instructions with additional strategies to achieve deeper processing and the question is how it compares to practice (Question 1: condition A = BC or condition A < BC?).

Heijltjes et al. (2014) also found that explicit CT instruction with practice and self-explanation prompts did not lead to better performance on practiced tasks compared to explicit instructions and practice without such prompts, but did improve performance on not-practiced tasks on an immediate test. We hypothesize that this finding would not be replicated when not practiced tasks are not instructed (Hypothesis 2: B = C on both practiced and not-practiced tasks).

An additional aim of this study was to replicate Heijltjes et al.'s (2014) exploratory findings regarding the role of students' dispositions in reasoning and learning to reason. That is, two instruments to measure thinking dispositions, 'actively open-minded thinking' (AOT; i.e., the active search for evidence against one's own beliefs, plans, or goals and the ability to weigh available evidence fairly; Baron, 2008) and Need for Cognition (NFC; i.e., "the tendency for an individual to engage in and enjoy thinking"; Cacioppo, Petty, & Kao. 1984, p. 306) have been shown to predict performance on a variety of tasks associated with rational thinking (Sá, Kelley, Ho, & Stanovich, 2005; Stanovich & West, 1997, 2007; West et al., 2008). The study by Heijltjes et al. (2014) confirmed that higher scores on these dispositions were associated with higher reasoning scores (which we expect to replicate: Hypothesis 3a). However, it also showed that scores on dispositions did not interact with effects of instructional conditions (i.e., instructions

**Table 1**
*Overview of the Study Design with Sequence of Events.*

| | | | | | | |
|---|---|---|---|---|---|---|
| | Pre-test reasoning skills test and disposition tests (i.e., AOT and NFC) ($N$ = 152) ↓ | | | | | |
| Conditions | Group A ($n$ = 25) | Group B ($n$ = 26) | Group C ($n$ = 25) | Group D ($n$ = 25) | Group E ($n$ = 25) | Group F ($n$ = 26) |
| | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| Intructions | CT-instruction text about 2 out of 4 categories of tasks (i.e., Wason selection and Conjunction task) + Reading CT-instruction text twice and underline and encircle key-elements | CT-instruction text about 2 out of 4 categories of tasks (i.e., Wason selection and Conjunction task) + CT-Practiced on CT instructed tasks | CT-instruction text about 2 out of 4 categories of tasks (i.e., Wason selection and Conjunction task) + CT-Practiced on CT instructed tasks + Self-explanation prompts on instructed plus practiced tasks | Unrelated text + Reading text twice and underline and encircle key-elements. | Unrelated text + CT-Practiced tasks On 2 out of 4 categories of tasks (i.e., Wason selection and Conjunction task) | Unrelated text + CT-Practiced tasks On 2 out of 4 categories of tasks (i.e., Wason selection and Conjunction + Self-explanation prompts on CT practiced tasks |
| | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| | Post-test reasoning skills test | | | | | |

*Note.* AOT = Actively Open-minded thinking, NFC = Need for Cognition



Contents

were effective for students regardless of their scores on dispositions; which we expect to replicate: Hypothesis 3b).

Finally, this study aimed to explore the relationship between invested mental effort and confidence in reasoning tasks and whether this would change as a consequence of instruction. Both confidence in responses on reasoning tasks and mental effort invested in those tasks seem to be associated with the extent to which people engage in Type 1 and Type 2 reasoning (Evans, 2012; Thompson, 2009). Type 1 thinking is automatic, imposing little if any load on working memory, while Type 2 thinking is deliberate and imposes high working memory load (Evans, 2008; 2012), as a consequence, one might expect that engaging in Type 2 thinking would lead to higher investment of mental effort, which is an index of the actual cognitive load experienced (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). As for confidence, it has been found that the fluency with which an answer comes to mind can lead to a 'feeling of rightness', which determines to what extent (more effortful) Type 2 reasoning is subsequently engaged in (Thompson, 2009). For example, participants who had a weak feeling of rightness about an initial judgment spent more time on reaching an answer, and were more likely to change their answer than participants with a strong feeling of rightness (Thompson, Prowse Turner, & Pennycook, 2011). Because people tend to engage in Type 1 thinking on reasoning tasks, it can be expected that prior to instruction, there will hardly be a relationship between performance and confidence (Griffin & Tversky, 2002; Hypothesis 4a), whereas confidence and effort are likely to be negatively related (Hypothesis 4b). Another interesting question, however, is whether levels of confidence and invested mental effort would change as a function of instruction. Students in the instruction conditions have learned how to reason about such tasks, so on the posttest, in the trained conditions (A, B, and C) performance on the trained tasks would be expected to become better while confidence in that performance might remain the same or even become higher and effort invested might remain the same or even become lower. On not-trained tasks, in contrast, students may realize from the training that they should not respond in an automatic manner, but they have not learned how to reason about these specific tasks. So one might expect them to engage in Type 2 reasoning, and although it is not clear whether this will improve performance, one would expect this to be associated with a decrease of confidence and an increase of effort on those not-trained tasks (Question 2).

Contents

# Method

## Participants and Design

Participants were 152 Economics students of a Dutch University of Applied Sciences (age *M* = 18.9, *SD* = 1.49; 119 males). Participants were randomly assigned to one of six conditions: A) CT-instruction text (*n* = 25), B) CT-instruction text combined with CT-practice (*n* = 26), C) CT-instruction text combined with CT-practice and self-explanation prompts (*n* = 25), D) unrelated text (*n* = 25), E) unrelated text plus CT-practice (*n* = 25), and F) unrelated text plus CT-practice and self-explanation prompts (*n* = 26).

## Materials

**Disposition tests.** We used a Dutch translation of the 41-item Actively Open-minded Thinking test (AOT; Stanovich & West, 2007) and the 18-item (short form) Need for Cognition questionnaire (NFC; Cacioppo et al., 1984) to measure CT dispositions. Both consist of items requiring a response on a 6-point Likert scale (AOT: as in Stanovich & West, 2007; NFC: as in West et al., 2008; note that Cacioppo et al., 1984 used a 9-point scale) ranging from (1) strongly disagree to (6) strongly agree. Scores on the items are summed for AOT and for NFC separately (after reverse scoring items that are formulated negatively). Higher scores on AOT characterize a greater tendency toward open-minded thinking. The reliability of the AOT was good: the Cronbach's alpha was .70. Higher scores on the NFC represent a greater tendency to engage in and enjoy effortful cognitive endeavors. Reliability of the NFC was also good: Cronbach's alpha was .80.

**CT skills tests***.* The CT skills tests consisted of eight tasks, across four categories: 1) *base-rate tasks*, which measure to what extent people are influenced by personal experience, vividly presented evidence from a single case, or prior beliefs in favor of more representative statistical evidence or prior probability (adapted from De Neys & Glumicic, 2008; Fong, Krantz, & Nisbett, 1986), 2) *conjunction tasks* (adapted from Tversky & Kahneman, 1983), that measure to what extent people neglect a fundamental qualitative law of probability, that is, the conjunction rule ($P$(A&B) ≤ $P$(B)) that states that a conjunction cannot be more probable than one of its ingredients, 3) *Wason selection tasks* (adapted from Stanovich, 2009; Wason & Shapiro, 1971), that measure the tendency to verify rules rather than to falsify them and the tendency to consider only that information as relevant which matches the lexical content of the statement regardless of its logical significance, and 4) *syllogistic reasoning tasks* adapted from Evans, 2003; Markovits & Nantel, 1989; Sá, West, & Stanovich, 1999) that examine the tendency to evaluate the logical validity of arguments on the basis of one's prior beliefs about the truth of conclusions. The content of the surface features (cover stories) of the tasks was adapted to the interests of students. A multiple-choice format was used with four answer options (the correct answer based upon rational reasoning strategies and incorrect answers related to biased reasoning). The tasks in the pre-test and post-test were structurally equivalent, but surface features (cover stories) differed.

**CT text / unrelated text.** The CT instructions (conditions ABC) consisted of an explanation of two of four categories of tasks seen in the pre-test: Wason selection and conjunction. The written text and graphics demonstrated the reasoning process step-by-step, going over all possible -right and wrong- answers. Examples were used in the demonstration referring to the tasks seen in the pre-test.

The unrelated text (conditions DEF) was a newspaper article approximately equal in length, about how advertising can be matched with a person's mood; participants were instructed to read this text attentively.

To equalize time on task with the practice conditions, conditions A and D were instructed to read the text twice and to underline and encircle the essential parts of the text, which, for the CT-instruction condition, could be a useful exercise.

**CT practice.** After reading the CT instruction text or the unrelated text, participants in condition B, C, E and F were exposed to a business-case from an economics course containing a description of a coffee manufacturer who had to decide about marketing, quality control, extending the assortment, and the sustainability measures. Then two Wason selection tasks and two conjunction tasks were practiced, these had a similar format as the tasks used in the pre-tests, but the cover stories of the practice tasks were derived from the business case.

**Self-explanation prompts.** During practice with the business case tasks, participants in condition C and F received prompts after each task to explain how the answer was obtained: 'Provide an explanation of how you've come to the answer, using keywords'.

**Mental effort.** Invested mental effort was measured with a 9-point subjective rating scale ranging from (1) very, very low effort to (9) very, very high effort (Paas, 1992).

**Confidence.** Participants rated their confidence in their response to each task by answering the question 'how certain are you that your response is correct?' on a 6-point rating scale ranging from (1) very uncertain to (6) very certain.

## Procedure

The experiment was run in 8 sessions in a computer room with 12 to 24 participants per session. All conditions were represented in each session and participants were randomly assigned to one. The pre-test and post-test materials were delivered in a computer-based environment that was created for this experiment and participants could work at their own pace. They first completed the pre-test and dispositions tests. After each task participants rated their invested mental effort and confidence. Subsequently, all participants entered the instruction phase, in which they first read the text, about CT or an unrelated topic, and then engaged in practice problems with or without prompts, depending on their assigned condition. At the end of the instruction phase participants were asked to rate the mental effort they invested in the instruction phase as a whole. Time-on-task was logged during instruction. After the instruction phase, participants performed a post-test comparable to the pre-test and again rated their invested mental effort and confidence after each task.

### Data Analysis

For each correct answer on the CT skills pre-test and post-test 1 point was assigned, resulting in a maximum score of 8 for each test. Scores on 4 trained and 4 not-trained tasks were computed in a similar manner. Difference scores between pre-test and post-test for trained and not trained tasks were computed for reasoning scores, invested mental effort and confidence in reasoning by subtracting the pre-tests scores from the post-test scores.

Ratings on the AOT and the NFC were summed after reverse scoring negatively formulated items, resulting in a maximum score on the AOT of 246 and on the NFC of 108.

Reported measures of effect size are $R^2$ and partial eta-squared (with 0.01, 0.06, and 0.14 signifying small, medium and large effects; Howell, 2012), and Cohen's *d* (with 0.20, 0.50, 0.80 signifying small, medium and large effects; Cohen, 1992).

## Results

For all analyses, a significance level of .05 was used. Random assignment of participants to conditions had been successful: Chi-square tests showed no significant diffences among conditions in educational background, $\chi^2(15) = 11.31$, $p = .730$ or gender $\chi^2(5) = 4.43$, $p = .490$, and multivariate analysis showed no differences in CT dispositions (AOT: $F(5, 146) = 0.63$, $p = .676$, $\eta_p^2 = .02$, NFC: $F(5, 146) = 1.19$, $p = .315$, $\eta_p^2 = .04$), pre-test performance $F(5, 146) = 1.51$, $p = .189$, $\eta_p^2 = .05$, mental effort invested in the pre-test $F(5, 146) = .68$, $p = .637$, $\eta_p^2 = .02$, or confidence on the pre-test $F(5, 146) = .65$, $p = .665$, $\eta_p^2 = .02$

Time was not accurately logged for all participants (group A, $n = 1$; group B, $n = 5$; group C, $n = 2$; group D, $n = 1$; group E, $n = 8$; group F, $n = 4$); their missing values were replaced with Multiple Imputation. Within IBM SPSS version 21, we used Markov Chain Monte Carlo (MCMC) Fully Conditional Specification (FCS) as imputation method, with predictive mean matching for the imputation of quantitative variables. This Bayesian method is suitable for data with an arbitrary pattern of missing values. A total of 10 completed datasets was generated, each of which being a combination of observed values (which are of course the same for all M datasets) and of imputed values (with probably different values for the M datasets). For each missing value (i.e., a particular case on a particular variable) we used the average value across the 10 completed datasets as value for imputation.

ANOVA on time on task during instructions revealed that time on tasks differed among conditions, $F(5, 146) = 6.71$, $p < .001$, $\eta_p^2 = .19$, indicating a medium effect; Post-hoc tests showed that condition F spent significantly more time on instructions than group A, B, and D (all *p*s < .001). Multivariate analysis revealed however, that time on tasks during instruction was not a mediating factor on post-test trained tasks, $F(1, 150) = .23$, $p = .630$, $\eta_p^2 = .00$, or not-trained tasks $F(1, 150) = .16$, $p = .687$, $\eta_p^2 = .00$).

## Reasoning Performance, Mental Effort and Confidence in Reasoning

Table 2 provides the means and standard deviations of CT test scores, invested mental effort and confidence in reasoning in each condition on trained and not-trained tasks on the pre-test and the post-tests. A multivariate analysis of covariance (MANCOVA) on difference scores (post-test minus pre-test) on trained and not trained tasks for reasoning performance, invested mental effort and confidence, with instruction conditions as between subjects factor and AOT and NFC as covariate revealed a significant effect of instruction conditions, Pillai's Trace, $F(30, 715) = 4.31$, $p < .001$, $\eta_p^2 = .15$. As can be seen in Table 3 only reasoning scores on trained tasks and invested mental effort on not trained tasks were significant.

ANOVA contrast tests on reasoning scores (see Table 4) of trained tasks revealed that participants in the CT-instruction conditions (ABC) had higher difference scores (i.e., learned more) than participants who did not received CT instructions (DEF). Other comparisons on reasoning scores between conditions ABC versus conditions DEF, or between condition A versus BC, between condition B versus C, or between condition D versus EF were not significant.

ANOVA contrast test on invested mental effort scores (see Table 4) revealed a significant higher difference score in invested mental effort on not trained tasks in the CT-instruction conditions (ABC) than the not CT-instructed conditions (DEF). Other comparisons on invested mental effort between conditions ABC versus conditions DEF,

**Table 2**
*Mean Scores and Standard Deviations (in Parentheses) on Trained and Not-trained tasks on the Pre-test and Post-test on Reasoning Scores, Invested Mental Effort Scores and Confidence Scores.*

| | Instructional conditions | | | | | |
|---|---|---|---|---|---|---|
| | A (*n* = 25) | B (*n* = 26) | C (*n* = 25) | D (*n* = 25) | E (*n* = 25) | F (*n* = 26) |
| Reasoning scores (max. score = 4) | | | | | | |
| Trained tasks pre-test | 1.32 (0.80) | 1.27 (0.87) | 1.28 (0.94) | 1.28 (0.79) | 0.88 (0.60) | 1.23 (0.86) |
| Trained tasks post-test | 3.52 (0.71) | 3.00 (1.20) | 3.12 (1.05) | 0.76 (0.78) | 0.64 (1.04) | 0.85 (0.94) |
| Not-trained tasks pre-test | 1.48 (1.05) | 1.15 (0.92) | 1.36 (0.86) | 1.68 (1.07) | 1.28 (0.94) | 1.00 (0.80) |
| Not-trained tasks post-test | 1.40 (0.96) | 1.46 (0.81) | 1.60 (0.76) | 1.44 (1.04) | 1.60 (1.12) | 1.00 (0.75) |
| Invested mental effort (max. score = 9) | | | | | | |
| Trained tasks pre-test | 3.45 (0.98) | 3.52 (1.15) | 2.99 (1.06) | 3.23 (1.29) | 3.48 (1.12) | 3.40 (0.91) |
| Trained tasks post-test | 2.86 (1.13) | 3.17 (1.50) | 2.89 (1.48) | 3.27 (1.32) | 3.02 (1.06) | 3.25 (1.30) |
| Not-trained tasks pre-test | 3.22 (1.02) | 3.32 (1.05) | 3.09 (1.21) | 3.16 (1.33) | 3.47 (1.19) | 3.44 (1.07) |
| Not-trained tasks post-test | 3.66 (0.85) | 4.09 (0.83) | 3.79 (1.17) | 3.21 (1.34) | 3.71 (0.99) | 3.25 (1.09) |
| Confidence (max. score = 6) | | | | | | |
| Trained tasks pre-test | 4.67 (0.75) | 4.65 (0.67) | 4.64 (0.81) | 4.54 (0.91) | 4.55 (0.63) | 4.50 (0.78) |
| Trained tasks post-test | 5,07 (0.64) | 4.56 (1.01) | 4.77 (1.00) | 4.67 (0.87) | 4.72 (0.74) | 4.43 (0.73) |
| Not-trained tasks pre-test | 4.84 (0.83) | 4.80 (0.66) | 4.86 (0.59) | 4.97 (0.74) | 4.79 (0.66) | 4.49 (0.85) |
| Not-trained tasks post-test | 4.61 (0.66) | 4.27 (0.73) | 4.32 (0.77) | 4.74 (0.84) | 4.47 (0.67) | 4.49 (0.74) |

*Note.* Conditions: A: critical thinking instruction; B: critical thinking instruction + practice; C: critical thinking instruction + practice + self-explanation prompts; D: unrelated text (control); E: unrelated text + practice; F: unrelated text + practice + self-explanation prompts.


Contents

or between condition A versus BC, between condition B versus C, or between condition D versus EF were not significant.

Contrasts on confidence scores (see Table 4) revealed lower difference scores on not-trained tasks in the CT-instruction conditions (ABC) than the not CT-instructed conditions (DEF), however it should be noted that the overall MANCOVA showed no significant effect of instruction conditions on confidence. Other comparisons on confidemce scores between conditions ABC versus conditions DEF, or between condition A versus BC, between condition B versus C, or between condition D versus EF were not significant.

To investigate relations between effort and confidence, bivariate correlation analyses were conducted (see Table 5), showing no significant correlation between confidence scores and performance on pre-test tasks and confidence scores and performance on not-trained post-test tasks. Only performance on post-test trained tasks showed a small to medium positive correlation with confidence, $r(152) = .200$, $p = .013$, $R^2 = .04$.

There was no significant correlation between invested mental effort and pre-test performance and a small to medium negative correlation with performance on post-test trained tasks, $r(152) = -.179$, $p = .028$, $R^2 = -.03$ (Table 5). As expected, there were large significant negative correlations between invested mental effort and confidence on both pre-test and post-test (all *ps* < .003).

## Dispositions

To investigate whether higher scores on dispositions (i.e., AOT and NFC) resulted in higher reasoning performance, bivariate correlation analyses were conducted (see

**Table 3**

*Reasoning Scores, Mental Effort and Confidence of Trained and Not-trained tasks as a Function of Instruction Condition*

| Variables[1] | *F* | *df* | *Error df* | *p* | $\eta_p^2$ |
|---|---|---|---|---|---|
| Reasoning scores on trained tasks | 38.45 | 5 | 144 | < .001 | .57 |
| Reasoning scores on not-trained tasks | 1.13 | 5 | 144 | .349 | .04 |
| Invested mental effort trained reasoning tasks | .84 | 5 | 144 | .522 | .03 |
| Invested mental effort not trained reasoning tasks | 3.38 | 5 | 144 | .006 | .11 |
| Confidence trained reasoning tasks | 1.41 | 5 | 144 | .225 | .05 |
| Confidence not trained reasoning tasks | 1.97 | 5 | 144 | .087 | .06 |

*Note.* [1]Scores were obtained by subtracting pre-test scores from post-test scores.

**Table 4**

*Contrasts between Conditions on Reasoning Scores, Mental Effort Scores and Confidence Scores on Trained and Not-trained tasks*

| | *Reasoning scores*[1] | | | | *Invested mental effort scores*[1] | | | | *Confidence scores*[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *t* | *df* | *p* | *d* | *t* | *df* | *p* | *d* | *t* | *df* | *p* | *d* |
| Scores on trained tasks | | | | | | | | | | | | |
| Condition ABC vs DEF | 13.77 | 146 | < .001 | 2.28 | - .84 | 146 | .403 | - 0.14 | .58 | 146 | .563 | 0.10 |
| Condition A vs BC | - .82 | 146 | .411 | - 0.14 | 1.25 | 146 | .212 | 0.21 | .45 | 146 | .652 | 0.07 |
| Condition B vs C | 1.65 | 146 | .102 | 0.27 | -1.33 | 146 | .187 | - 0.22 | 2.20 | 146 | .029 | 0.36 |
| Condition D vs EF | - .38 | 146 | .706 | - 0.06 | - .78 | 146 | .439 | - 0.13 | - 1.13 | 146 | .259 | -0.19 |
| Scores on not-trained tasks | | | | | | | | | | | | |
| Condition ABC vs DEF | .79 | 146 | .430 | 0.13 | 3.61 | 146 | < .001 | 0.60 | - 2.10 | 146 | .038 | - 0.35 |
| Condition A vs BC | -1.63 | 146 | .106 | - 0.27 | .10 | 146 | .918 | 0.02 | - .39 | 146 | .696 | - 0.06 |
| Condition B vs C | - 1.44 | 146 | .152 | - 0.24 | - 1.17 | 146 | .244 | - 0.19 | 1.70 | 146 | .091 | 0.28 |
| Condition D vs EF | .24 | 146 | .811 | 0.04 | .24 | 146 | .811 | 0.04 | .05 | 146 | .957 | 0.01 |

*Note.* [1]Scores were obtained by subtracting pre-test scores from post-test scores. Conditions: A: critical thinking instruction; B: critical thinking instruction + practice; C: critical thinking instruction + practice + self-explanation prompts; D: unrelated text (control); E: unrelated text + practice; F: unrelated text + practice + self-explanation prompts.

Table 4), revealing a significant relationship between AOT and pre-test reasoning performance and between AOT and post-test not-trained reasoning performance. NFC was not related to either pre-test reasoning performance, or post-test reasoning performance. To investigate potential interaction effects between instruction conditions and dispositions, a MANCOVA on reasoning scores, invested mental and confidence in reasoning, with conditions as between subject factor and dispositions (i.e., AOT and NFC) as covariates, shows no interaction effects between dispositions and instruction conditions (Pillai's Trace, AOT: $F(30, 665) = 1.06$, $p = .378$, $\eta_p^2 = .05$); NFC: $F(30, 665) = .75$, $p = .834$, $\eta_p^2 = .03$)

## Discussion

The main aim of this study was to investigate which training method would be most effective for improving first year economics students' reasoning skills. Results showed that explicit CT-instructions improved reasoning performance compared to no (i.e., mock) instructions (Hypothesis 1a). It seems that in order to overrule Type 1 reasoning, knowledge of reasoning principles and strategies has to be acquired. This is in line with the notion that biased thinking may be due to a 'mindware gap' (e.g., Stanovich, 2011).

Practicing tasks without instructions, either with or without self-explanation prompts, was not effective for learning (i.e., condition D = EF; Hypothesis 1b). Interestingly, combined with instructions, practice had no additive effect on learning compared to instructions only (i.e., condition A = BC; Question 1). It should be noted that an important difference with the Heijltjes et al. (2014) study, was that we attempted to equalize time on task between the conditions that did and did not receive practice tasks, by instructing participants in condition A to read the CT-instruction twice while underlining and encircling key statements the second time. Even though this method has been classified as relatively ineffective for learning compared to practicing methods (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), it seems to have led to deeper processing in our study. This suggests that it might not be the practice activity per se which affected learning in the Heijltjes et al. study, but the amount of time students engaged processing the central concepts and principles (Lee & Anderson, 2012). However, this finding would need to be replicated in future studies before we can conclude this with any certainty.

Prompting self-explanations during practice did not have a beneficial effect on either trained or not-trained tasks in this study (condition AB = C; Hypothesis 2). Thus, when tasks that were not practiced, had not been instructed either, there was no 'transfer' effect of prompting self-explanation, presumably because participants had no prior knowledge to base their explanations upon. We suppose that this was also the

**Table 5**
*Correlations between Dispositions (AOT and NFC), Performance, Confidence Rating and Invested Mental Effort Rating on Pre-test and Post-test Trained and Not-trained Tasks*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Actively Open-minded Thinking (AOT) | - | | | | | | | | | |
| 2. Need for Cognition (NFC) | .193 | - | | | | | | | | |
| 3. Pre-test tasks | .225 ** | .107 | - | | | | | | | |
| 4. Pre-test mental effort | - .021 | .051 | .056 | - | | | | | | |
| 5. Pre-test confidence | .060 | .051 | - .082 | - .647*** | - | | | | | |
| 6. Post-test trained tasks | .143 | .107 | .266 ** | - .065 | .061 | - | | | | |
| 7. Post-test mental effort trained tasks | - .097 | - .077 | - .036 | .578*** | - .468*** | - .179* | - | | | |
| 8. Post-test confidence trained tasks | .078 | .070 | .090 | - .453*** | .646*** | .200* | - .690*** | - | | |
| 9. Post-test not-trained tasks | .246 ** | - .073 | .374** | - .011 | .038 | .211* | - .067 | .058 | - | |
| 10. Post-test mental effort not- trained tasks | - .056 | - .157 | - .014 | .523*** | - .342*** | .096 | .540*** | - .332*** | - .028 | - |
| 11. Post-test confidence not-trained tasks | .045 | .180* | .113 | - .308*** | .536*** | - .042 | - .308*** | .501*** | .078 | - .654*** |

*Note*. *** Significant at the 0.001 level, **Significant at the .001 level, *Significant at the 0.05 level.



Contents

reason why self-explanation prompts were not effective for improving performance in the no CT-instruction conditions (condition i.e., DE = F). One limitation of this study, which future research should address, is whether particpants who were prompted to self-explain would perform better when the reasoning test would involve recall (i.e., open answer items) rather than recognition (i.e., multiple choice items), as recognition tests may not require the knowledge elaboration that one can expect self-explanation to result in.

As for our second aim, we replicated the findings by Heijltjes et al. (2014) that participants with higher scores on AOT showed better pre-test and post-test performance than participants with lower scores on AOT (Hypothesis 3a). However, in contrast to their findings, we found no correlation between NFC and reasoning performance on pre-test and post-test. Possibly, because NFC is strongly associated with intrinsic cognitive motivation (Cacioppo et al., 1996), the external regulation in the experimental setting overruled potential effects of NFC. As expected, there were no interaction effects between instruction conditions and dispositions (Hypothesis 3b), which indicates that instructions were equally effective for all participants regardless of their disposition scores.

As expected, reasoning performance and confidence scores were poorly related on pre-test (Hypothesis 4a), and confidence and effort were negatively related: the more certain participants were of their answers, the less mental effort they invested, not only on the pre-test (Hypothesis 4b) but even after training on post-test (Question 2). This seems in line with the 'fluency principle' (Thompson, 2009). On post-test trained tasks, confidence scores and reasoning performance were positively related (Question 2). Participants also seemed to become more cautious from pre-test to post-test, that is, participants in conditions ABC reported in contrast to participants in condition DEF, lower confidence on not-trained tasks, and equal confidence on trained tasks, even though for participants in condition ABC performance improved. In other words, participants in these conditions seemed to have become better calibrated; although we cannot test this assumption based on our data. Because appropriate confidence levels are associated with the quality of subsequent decisions (Paese & Smiezek, 1991), rational choices (Griffin & Tversky, 2002), and the extent to which people engage in Type 2 reasoning (Thompson, 2009), this is an important issue to explore further in future research.

On trained tasks participants had to invest equal mental effort on the post-test, indicating that in the CT instructed conditions (condition ABC) a better performance was reached with equal investment of effort. This implies that students in the instruction conditions allocated their cognitive resources more effectively (Hoffman & Schraw, 2010; Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). More interestingly, on not-trained tasks, participants in the instruction conditions did not enhance their reasoning performances but they show an increase in invested mental effort, especially in the CT instruction condition (ABC). This might be an indication that the instructions stimulated participants' not to respond automatically on new, not-trained tasks (i.e., engage in Type 2 instead of Type 1 thinking; Evans, 2011). However, given that they still lack the appropriate mindware to be successful in terms of the correct reasoning responses, this effort is not associated with better performance.

This study has some limitations. Despite attempts to equalize time on task, there was still a difference between conditions, although this cannot explain the beneficial effects of instructions on performance (since it was condition F that had the highest time on task). Another limitation is that because first year students were involved, the results might not generalize to intermediate or advanced students.

In sum, this study shows that first year economics students could enhance their reasoning performance by means of a short and relatively straightforward instructional intervention that is easy to implement in higher professional education. It would be worthwhile to explore in future studies how instructions can be integrated into the curriculum over time, in order to ingrain these skills for long term use in educational and professional contexts.

# 5 Summary and General Discussion

As outlined in the introduction, contemporary higher education has the challenging task of preparing students for the 21st century world, in which critical thinking is a crucial competence (Pellegrino & Hilton, 2012). Especially in higher education programs that seek to prepare students for working in dynamical and complex professional environments, such as economics, law, or medicine, it is imperative that erroneous decisions as a result of biased reasoning are prevented, because of the severe consequences they can have in terms of financial losses, emotional damage, and even human lives (e.g., Chapman & Elstein, 2000; De Bondt & Thaler, 2002; Tversky & Kahneman, 1974).

The question is, however, which instructional methods are most effective for supporting the acquisition of critical thinking skills in higher professional education, in particular with regard to unbiased reasoning. Therefore, the aim of the three empirical studies presented in this dissertation was to test the impact of different critical thinking instructions on one essential aspect of economics students' critical thinking, that is, the ability to enagage in unbiased reasoning. The first and main aim was to investigate the effects of general explicit critical thinking instructions (as compared to implicit instruction or no instruction) and the effcets of practice in a domain context (with or without prompts to foster deeper reasoning), on the acquisition of reasoning skills. The second aim was to explore the role of thinking dispositions and whether they interact with the effects of instructions. The third aim was to explore the role of confidence and mental effort in reasoning prior to and after instruction. All three experiments were conducted in departments of Business and Economics at a Dutch University of Applied Sciences.

In this chapter the main findings of the studies are summarized and discussed, along with methodological considerations, implications for educational practice, and suggestions for future research.

## Aim 1: Investigating Effects of Critical Thinking Instructions, Practice, and Prompts

The main aim of the studies reported here, that is, investigating effects of critical thinking instructions, was addressed in all three experiments. The study in Chapter 2 showed that implicit instruction by means of engaging in a regular course about argumentation and negotiation skills was not sufficient for fostering performance on the reasoning tasks. Only explicit general instructions combined with practice on a domain-specific

Contents

case were effective; there was no improvement on tasks that had been instructed, but not practiced, or on tasks that had been practiced, but not instructed. There was no effect of prompting during practice; neither self-explanation prompts, nor activation prompts enhanced performance on practiced tasks compared to the condition that received explicit instructions and practice without those prompts.

The study reported in Chapter 3 replicated these findings. Again, only explicit, general instructions combined with practice on a domain-specific case led to higher post-test reasoning performance and there was no improvement on tasks that had been instructed, but not practiced, or on tasks that had been practiced, but not instructed. The only exception to that was that the condition that was prompted to self-explain during practice, showed better performance on the tasks that were instructed but not practiced on the immediate test. However, this effect had disappeared after a three week delay. Activation prompts did not improve critical thinking performance. This study did show that the effect of the combination of instructions plus practice was stable over a three week delay.

Although the studies in Chapter 2 and 3 showed that participants who practiced tasks after explicit instruction improved their reasoning performance, this raised the question of whether this effect was due to the specific qualities of practice or because participants in those conditions spent more time on the learning phase. Moreover, in the study presented in Chapter 3 it was found that adding self-explanation prompts during practice enhanced immediate post-test performance on not-practiced tasks that had been instructed, however it was unclear whether this effect would also arise if no prior instructions were provided.

Therefore, in the study presented in Chapter 4 the control condition included instructions that were expected to results in additional meaningful time-on-task, and the question was addressed of whether self-explaining during practice would also aid transfer without any prior instructions about the tasks. The study showed that only students who received explicit critical thinking instructions improved reasoning performance compared to no (i.e., mock) instructions, but again only on tasks that were trained (i.e. instructed or instructed plus practiced). The results showed that practice had no additive effect on learning compared to an alternative deep-processing approach of the instructions. Prompting self-explanations during practice did not have a beneficial effect on trained tasks as in Chapters 2 and 3, nor on not-trained tasks. Thus, when tasks that were not practiced, had not been instructed either, we found no 'transfer' effect of prompting self-explanation.

In sum, the studies presented in this dissertation show that explicit critical thinking instructions are necessary but probably not sufficient to improve unbiased reasoning; critical thinking instructions were only effective in Chapters 2 and 3 when combined

Contents

with practicing tasks, although the study in Chapter 4 showed that engaging in actively processing textual instructions can be as effective as practice. We found no evidence of transfer to tasks that were instructed but not practiced (Chapters 2 and 3) or tasks that were not trained at all (Chapter 4). Only prompting self-explanations seemed to evoke a kind of 'transfer' to tasks that were instructed but not practiced (Chapter 3), but this was only found at an immediate and not on the delayed test. Moreover, this effect was not found when tasks had not been instructed first (Chapter 4).

The results confirm and expand previous research in several ways. In line with research on other aspects of critical thinking (e.g., Arum & Roksa, 2011; Bangert-Drowns & Bankert, 1990; Marin & Halpern, 2011), our findings show that unbiased reasoning does not develop spontaneously as a side effect of education, but that it has to be taught explicitly. Moreover, in line with research on other aspects of critical thinking (e.g., Abrami et al., 2008; Angeli & Valanides, 2009), our findings show that also for improving reasoning, explicit general instruction combined with practicing tasks in a domain-specific context is beneficial (Chapters 2, 3 and 4), although an alternative strategy for deep processing of the general instructions seemed as effective in Chapter 4. This is in line with the dual processing point of view, which states that in order to overrule Type 1 processes, the appropriate 'mindware' has to be available, that is, knowledge of reasoning principles and strategies has to be acquired and present in mind (Stanovich, 2011). Apparently, viewing a single explicit video-based critical thinking instruction was not sufficient to acquire the appropriate mindware to overrule Type 1 processes. However, also practicing the tasks in an economics context (i.e., on a business case; Chapter 2, 3, 4), or processing a written instructional text more deeply by re-reading and underlining and encircling key statements, presumably enabled students to mentally organize and reorganize the knowledge they gained from the instructions better, which facilitated retrieval from memory in subsequent reasoning tasks.

It should also be noticed that it is unclear whether the difference in findings regarding the instruction only condition in the studies in Chapters 3 and 4 is due to the fact that instead of a video instruction (which is transient) a written text was used (which always allows re-reading parts even if re-reading and underlining would not be encouraged), and that it is unclear whether the effectiveness of the rereading approach would also remain after a delay (as the effects of instructions of practice were shown too, in Chapter 3).

Prompting self-explanation was expected to foster the proper use of available knowledge (Lombroso, 2006; Roy & Chi, 2005), however those prompts did not foster performance on practiced reasoning tasks that had (Chapter 2 and 3) or had not been (Chapter 4) instructed. Probably critical thinking instruction combined with practicing tasks was sufficient to integrate and retrieve novel information, and as a consequence, self-explanations did not have additional effects on performance on the practiced tasks (but the answer format may have played a role here; see also the section on Methodological considerations). However, in line with previous research that showed the effect of self-explanation on transfer –for example on problem solving (Aleven & Koedinger, 2002) or on algorithmic tasks (Renkl, Stark, Gruber, & Mandl, 1998; Rittle-Johnson, 2006)- the findings from Chapter 3 suggest that on tasks that were instructed but not practiced, prompting self-explanations established a kind of transfer from practiced to not-practiced tasks compared to the control condition. However this effect should be interpreted with caution, given that the difference with the other instruction conditions was not significant, that the effect was no longer present on the delayed test, and that this was not found in the study presented in Chapter 2. It might be worthwhile to attempt to deepen instructional effects of self-explanation in future studies (see section Recommendations for future research).

Activation prompts, which were expected to be able to affect critical thinking performance by redirecting attention to relevant cues (Moutier, Angeard, & Houdé, 2002; Moutier & Houdé, 2003), did not lead to better test performance than instructions only (Chapter 2 and 3). This finding might indicate that instructions were sufficient to redirect attention and inhibit initial automatic responses during practice.

In sum, these findings imply that the improvements of students' unbiased reasoning relies on explicit instruction of general rules and reasoning principles combined with strategies for deeper processing, regarding which it has been shown that practice in a domain context is a suitable approach with lasting effects.

## Aim 2: Exploring the Impact of Thinking Dispositions

The second aim was to explore the role of thinking dispositions and whether they interact with the effects of instructions. All three experiments reported in this dissertation revealed that some students showed more biased reasoning than others, and that their scores on dispositions predicted their reasoning performance. However, the findings on dispositions were inconclusive. Dispositions were measured by both AOT (i.e., Actively Open-minded thinking; the active search for evidence against one's own beliefs, plans, or goals and the ability to weigh available evidence fairly; Baron, 2008), and NFC (i.e., Need for Cognition, "the tendency for an individual to engage in and enjoy thinking"; Cacioppo, Petty, & Kao, 1984, p. 30). Higher scores on AOT were consistently associated with better initial performance on reasoning tasks (Chapter 2, 3 and 4); however higher scores on NFC only predicted better initial reasoning performance in the study presented in Chapter 3. The results on AOT converge with findings in other studies (Macpherson & Stanovich, 2007;Sá, Kelley, Ho, & Stanovich, 2005; Stanovich

Contents

& West, 1997; West, Toplak, & Stanovich, 2008). Regarding NFC, it is possible that the external regulation provided by the experimental setting overrode potential effects of NFC, because NFC is strongly associated with intrinsic cognitive motivation (Cacioppo, Petty, Feinstein, & Jarvis 1996). Possibly, participants in the study reported in Chapter 3 might have experienced such external regulation to a lesser extent, because they were part-time continuing education students, in contrast to the students who participated in the studies in Chapter 2 and 4.

An even more interesting question regarding students' dispositions that was addressed was whether they would interact with the effects of instructional conditions (e.g., whether students with higher scores on dispositions would benefit more or less from the critical thinking instruction than students with lower scores on dispositions). None of the studies reported in this dissertation showed a significant interaction between instruction conditions and dispositions. Thus these explorative findings suggest that all students benefitted equally from the instructions, and that under conditions in which explicit instructions are provided, AOT and NFC do not influence learning. This seems to be good news for educators, although students with lower scores on dispositions might require additional attention or more training to get them on the same level as students with higher scores (because even though they benefit equally, they started at a lower level of performance).

### Aim 3: Exploring Invested Mental Effort and Confidence in Reasoning

The third aim of the studies reported here, was to explore the role of confidence and mental effort in reasoning prior to and after instruction. In Chapter 2 and 4, invested mental effort on the reasoning tests prior to and after instruction was investigated to explore potential effects of instructions (and practice) on the cognitive demands imposed by reasoning tasks. Two options were considered plausible: knowledge of strategies gained through instructions might reduce cognitive load imposed by the task, and therefore lead to less effort investment on the post-test than on the pre-test (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), or instructions might evoke engaging in more effortful Type 2 reasoning (Evans, 2011). In both studies, there were no differences in invested mental effort scores on the pre-test or post-test among instruction conditions. Given that the explicit critical thinking instruction conditions with additional practice or rereading assignment outperformed the other conditions, it is clear that even though they invested the same amount of effort, this effort was apparently invested in very different cognitive processes. In other words, instructional condition did affect the efficiency of performance in those conditions: a better performance was reached with equal investment of effort during practice and after instruction (Hoffman & Schraw, 2010; Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008).

Interestingly the study presented in Chapter 4 revealed that on not-trained tasks (which were neither instructed nor practiced), participants in the instruction conditions did not enhance their reasoning performances but they show an increase in invested mental effort. This might be an indication that having had instructions on other tasks stimulated participants' not to respond automatically on new, not-trained tasks (i.e., engage in Type 2 instead of Type 1 thinking; Evans, 2011). However, given that they still lacked the appropriate mindware to be succesful in terms of correct reasoning, this additional effort was not associated with better performance.

As can be inferred from the study in Chapter 4, confidence and invested mental effort were negatively related: the more certain participants were of their answers, the less mental effort they invested in determining their answer, not only on the pre-test but even after training on post-test. This seems in line with the 'fluency principle' (Thompson, 2009) which implies that the fluency with which an answer comes to mind can lead to a 'feeling of rightness', that subsequently determines the extent of engaging in more effortful reasoning. It can also be explained by the 'sufficiency principle' (Chen & Chaiken, 1999) suggesting that an individual exerts as much mental effort as necessary to reach a satisfying level of confidence, as people are 'economy minded'; they balance between minimizing their effort investment and feeling sufficiently confident. While reasoning performance and confidence scores were poorly related on pre-test, they were positively related on post-test trained tasks. Also, participants in the conditions that had received critical thinking instruction seemed to become more cautious from pre-test to post-test, that is, they reported lower confidence on not-trained tasks, and equal confidence on trained tasks, even though their performance improved on trained tasks. It therefore seems that those participants have become better calibrated, that is, better able to judge their own performance; although it was not possible to test this assumption based on the data. Because appropriate confidence levels are associated with the quality of subsequent decisions (Paese & Smiezek, 1991), rational choices (Griffin & Tversky, 2002), and the extent to which people engage in Type 2 reasoning (Thompson, 2009), this is an interesting finding that is worth exploring further in future research.

## Methodological Considerations

This section gives a reflection on the strengths and limitations of the studies presented in this dissertation. To start with the latter, because limitations of the individual studies have already been reported in each chapter, they will not be reiterated here; only some more general limitations will be discussed. First, regarding the generalization of findings, it should be kept in mind that the participant population in these studies was restricted to business and economics students. It can be argued that these students differ from other majors; for instance, Arum and Roksa (2011) found that business and economics students had the lowest scores on critical thinking and complex reasoning, compared to other college majors. Nevertheless, proneness to biased reasoning has been identified as a general phenomenon among lots of students in many disciplines (e.g., Tversky & Kahneman, 1983), so there is no reason to assume that a similar approach to instructing critical thinking would not be effective in other domains, but future research should establish that.

Another potential limitation is that in hindsight, the answer formats (i.e., multiple choice) may not have allowed us to measure whether participants really understood the reasoning behind the right answer. It could be argued that it would have been better to test students' recall in open answer reasoning tests and ask them to explain their answers rather than testing their recognition in multiple choice tests. This would, for instance, have enabled us to test effects of self-explanation prompts better, because recognition tests do not assess the degree of knowledge elaboration that one can expect self-explanation to result in.

Finally, a limitation of the studies presented here, is that the task types that were practiced or not practiced (Chapters 2 and 3) or trained and not trained (Chapter 4) were not counterbalanced. Although it seems unlikely that there were differences in task characteristics between tasks that were practiced/trained or not practiced/trained and we cannot be sure whether our results were affected by those differences .

A strength of the studies reported here is that they were conducted at the University of Applied Sciences, in students' own educational context, yet with experimental control. The study presented in Chapter 2 was even embedded within a course. By conducting the studies close to the educational practices of everyday, the scientific findings were directly shared in educational practice, and encouraged thinking about evidence-based strategies by educators. This is one of the advantages of 'moving educational science from lab settings into controlled field studies' as recently recommended by Roediger (2013).

## Implications for Educational Practice

The findings of this dissertation clearly indicate that educators can and should have an active role in teaching students how to avoid biased reasoning. Although John Dewey already argued that habits of reflective thoughts are "not a gift of nature" (Dewey, 1997, p. 28), this is still not widely acknowledged in educational practice. Unfortunately, many educators seem to believe that critical thinking skills develop spontaneously as a side effect of higher education or immersion into a discipline (Davies, 2013; Paul, 2005). In contrast to this believe, based on these studies it is recommended that critical thinking should be explicitly embedded in the higher education curricula. This is not only a responsibility of educators, but also of educational developers and advisors who are engaged in curriculum innovation and evaluation.

What complicates their task, however, is that it is not always clear what works best and why. The findings from this dissertation showed that students only learned when a combination of explicit general critical thinking instruction was combined with additional activities to process the instructions deeply. Moreover, instructing students about human tendencies to fall prey to cognitive bias necessitates educators to understand those biases and fallacies themselves, so critical thinking instruction would have to become an important issue in educators' professional development.

What is promising for educational practice, is that the studies in this dissertation showed that reasoning skills could already be enhanced through relatively straightforward and short instructional interventions, which are easy to implement in higher professional education and could easily be repeated within or across courses. The explorative findings regarding dispositions indicate that it were not only the higher achievers on thinking dispositions who learned. It was found that *all* learners in the explicit instruction conditions benefited from the instructions.

## Recommendations for Future Research

Although the studies in this dissertation have shown that reasoning performance can be improved by means of short interventions, it would be worthwhile to explore in future longitudinal studies how instructions can be integrated into the curriculum over time, in order to ingrain these skills for long term use in educational and professional contexts. For instance, would it be effective to expose students to explicit instructions and/or practice several times during a course? Repeating instructions throughout a course or even throughout an entire economics curriculum, might have much stronger effects, and might help economics and business students to learn to avoid biased reasoning when making decisions in dynamic and complex business environments. Another interesting question for future research to address is whether and under what conditions students would be able to transfer critical thinking skills to other types of tasks or courses. The studies presented here showed little evidence of transfer, but possibly students need to be explicitly prompted or reminded that they can use what they have learned about other tasks to reason about new tasks.

Contents

Future research might also further investigate why prompts seem to have no additional benefits for performance, and whether such benefits might occur under different circumstances, for example when students would be reminded prior to the test to ask themselves similar questions as they received during practice. Another question is whether particpants who were prompted to self-explain would perform better when the reasoning test would involve recall (i.e., open answer items) rather than recognition (i.e., multiple choice items). Recognition tests probably rely less on knowledge elaboration and therefore might not show effects of knowledge elaboration (e.g., Neilens, Handley, & Newstead, 2008). It might also be worthwhile to attempt to deepen instructional effects of self-explanation in future studies by explicitly teaching it as a meta-strategy or by providing feedback on self-explanations, as this might further enhance the effects of self-explanation on critical thinking and potentially on transfer to not-practiced tasks.

Finally, an interesting question for future research is whether students' dispositions can be improved by instruction. For example would repeatedly instructing students throughout a curriculum to think critically, also shape their beliefs, cognitive styles, goals, or epistemic values? Given the importance of critical thinking, the ultimate goal of education should be to deliver students who show their thinking abilities in educational and professional contexts without being prompted to do so.

# References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*, 1102-1134.

Aleven, A. W., & Koedinger, K. R. (2002). An effective meta-cognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147-179.

Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill-defined issues. *Learning and Instruction*, *19*, 322-334.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on colleges campuses.* Chicago: The University of Chicago Press.

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples. *Educational Psychological Review, 19*, 375–386.

Austin, Z., Gregory, P. A., & Chiu, S. (2008). Use of reflection-in-action and self-assessment to promote critical thinking among pharmacy students. *American Journal of Pharmaceutical Education, 72*, 1-8.

Bangert-Drowns, R. L., & Bankert, E. (1990). *Meta-analysis of explicit instruction for critical thinking.* Boston, MA: Paper presented at the annual meeting of the American Educational Research Association.

Baron, J. (2008). Actively open-minded thinking. In J. Baron (Ed.), *Thinking and Deciding* (pp. 199-232). New York: Cambridge University Press.

Billing, D. (2007). Teaching for transfer of core/key skills in higher education. *Higher Education*, *53*, 483-516.

Bok, D. (2006). *Our underachieving colleges: A candid look at how much students learn and why they should be learning more.* Princeton University Press.

Borg, J. R., & Borg, M. O. (2001). Teaching critical thinking in interdisciplinairy economics courses. *College Teaching, 49,* 20-25.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem-solving. *American Psychologist, 41*, 1078-1089.

Braun, N. M. (2004). Critical thinking in the business curriculum. *Journal of Education for Business*, 232-236.

de Bruin, W., Parker, A. M., & Fischoff, B. (2007). Individual differences in adult decion-making competence. *Journal of Personality and Social Psychology, 92*, 938-956.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197-253.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

Carrithers, D., Ling, T., & Bean, J. C. (2008). Messy problems and lay audiences: Teaching critical thinking within the finance curriculum. *Business Communication Quarterly, 71*, 152-170.

Celuch, K., & Slama, M. (1999). Teaching critical thinking skills for the 21st century. *Journal of Education for Business*, *3*, 134-139.

Chapman, G. B., & Elstein, A. S. (2000). Cognitive processes and biases in medical decision making. In G. B. Chapman, & F. A. Sonnenberg (Eds.), *Decision making in health care: Theory, psychology and applications* (pp. 183-210). Cambridge: Cambridge University Press.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken, & Y. Trope (Eds.), *The dual-process theories in social psychology* (pp. 73-96). New York: The Guilford Press.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18*, 293-328.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112*, 155-159.

Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing, 3*, 97-108..

Croskerry, P. (2003). Cognitive forcing strategies in clinical decision making. *Annals of Emergency Medicine, 41*, 110-120.

Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development, 32*, pp. 1-16.

De Bondt, W., & Thaler, R. H. (2002). Do analyst overact? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 678-685). New York: Cambridge University Press.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248-1299.

Dewey, J. (1997). *How we think.* New York: Dover Publications, Inc.

Dunlosky, J., Rawson, K. A., Marsh, J. E., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4-48.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*, 98-121.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*, 44-48.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*, 4-10.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454-459.

Evans, J. St. B. T. (2008). Dual Processing account of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*, 86-102.

Evans, J. St. B.T. (2012). Spot the difference: Distinguishing between two kinds of processing. *Mind and Society, 11*, 121-131.

Evans, J. St. B.T., Handley, S. J., & Harper, C. N. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Journal of Experimental Psychology, 54*, 935-958.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* Millbrae CA: The California Academic Press.

Facione, P. A. (2009). *Critical thinking: What it is and why it counts.* Retrieved from http://www.insightassessment.com/9articles%20WW.html

Flores, K. L., Matkin, G. S., Burbach, M. E., Quinn, C. E., & Harding, H. (2012). Deficient critical thinking skills among college graduates: Implications for leadership. *Educational Philosophy and Theory, 44*, 212-230.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effect of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253-292.

Gilovich, T., & Griffin, D. W. (2002). Introduction heuristics and biases: Then and now. In T. Gilovich, D. W. Griffin, & D. Kahnemen (Eds.), *Heuristic and biases: The psychology of intuitive judgment* (pp. 1-18). New York: Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002) (Eds.). *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Griffin, D. W., & Tversky, A. (2002). The weighing of evidence and the determinants of confidence. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 230-249). New York: Cambridge University Press.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*, 449-455.

Halpern, D. F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning, 80,* 69-74.

Harasym, P. H., Tsai, T. C., & Hemmati, P. (2008). Current trends in developing medical students' critical thinking abilities. *The Koashing Journal of Medical Sciences, 24*, 341-355.

Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning & Instruction, 29,* 31-42.

Helsdingen, A., Van Gog, T., & Van Merriënboer, J. J. G. (2011). Effects of practice schedule and critical thinking instruction on learning and transfer of a complex judgment task. *Journal of Educational Psychology, 103*, 383-398.

Hilbert, T. S., Schworm, S., & Renkl, A. (2004). Learning from worked-out examples; The transition from instructional explanations to self-explanation prompts. *Instructional design for effective and enjoyable computer-supported learning*, 184-192.

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist, 45*, 1-14.

Contents

Howell, D. C. (2012). *Statistical methods for psychology* (8th edition). Australia: Wadsworth.

Jones, A. (2007). Multiplicities or manna from heaven? Critical thinking and the disciplinary context. *Australian Journal of Education, 51*, 84-103.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 273-251.

Kahneman, D., & Lavallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science, 39*, 17-31.

Klaszynski, P. A. (2006). Learning, belief biases, and metacognition. *Journal of Cognition and Development, 7,* 295-300.

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology, 79*, 470-485.

Klebba, J. M., & Hamilton, J. G. (2007). Structured case analysis: Developing critical thinking skills in a marketing case course. *Journal of Marketing Education*, *29*, 132-139.

Knauff, M., Budeck, C., Wolf, A. G., & Hamburger, K. (2010). The illogicality of stock-brokers: Psychological experiments on the effects of prior knowledge and belief biases on logical reasoning in stock trading. *PloS one, 5*, 1-6 e13483.

Koehler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 686-715). New York: Cambridge University Press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*, 16-26.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment & decision making* (pp. 316-337). Malden, USA: Blackwell Publishing.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefits reasoning in everyday life. *Psychological Science, 1*, 362-370.

Lee, H.S., & Anderson, J. R. (2013). Student learning: What has instruction got to do with It? *Annual Review of Psychology, 64*, 445-469.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning. *American Psychologist, 43*, 431-442.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London: Sage.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*, 464-470.

Lunn, P. D. (2011). *The role of decision-making biases in Ireland's banking crisis.* Economic and Social Research Institute (ESRI).

Macpherson, R., & Stanovich, K. E. (2007). Cogitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences, 17*, 115-127.

Mamede, S., Van Gog, T., van den Berge, K., Rikers, R. M. J. P., van Saase, J. L.C. M., van Guldener, C., et al. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA, 304*, 1198-1203.

Marin, L. M., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6, 1-13.

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition, 17*, 11-17.

Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations concepts and procedures as pedagogical tools. *Journal of Experimental Child Psychology, 102*, 1-21.

McEldoon, K. L., Durkin, K. L., & Rittle-Johnson, B. (2012). Is self-explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology*, 1-18.

McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education, 26*, 3-29.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*, 10-12.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Moore, T. J. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development, 30*, 261-274.

Contents

Moutier, S., Angeard, N., & Houdé, O. (2002). Deductive reasoning and matching-bias inhibition training: Evidence from a debiasing paradigm. *Thinking and Reasoning, 8*, 205-224.

Moutier, S., & Houdé, O. (2003). Judgement under uncertainty and conjunction fallacy inhibition training. *Thinking and Reasoning, 9*, 185-201.

Neilens, H. L., Handley, S. J., & Newstead, S. E. (2008). Effects of training and instruction on analytic and belief-based reasoning processes. *Thinking & Reasoning, 15*, 37-68.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625-631.

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review, 9*, 114-128.

Nussbaum, M. C. (2006). Education and democratic citizenship: Capabilities and quality education. *Journal of Human Development, 7*, 385-395.

Paas, F. (1992). Training strategies for attaining transfer of problem solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429-434.

Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental efforts and performance measures. *Human factors, 35*, 737-743.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitieve load theory. *Educational Psychologist, 38*, 63-71.

Paese, P. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgments: Practice, effort, information and decision-making. *Organizational Behavior and Human Decision Processes, 48*, 100-130.

Paul, R. W. (1985). McPeck's mistakes. *Informal Logic, 7*, 35-43.

Paul, R. W. (1990). Critical thinking: What, why and how. In A. J. Binker (Ed.), *Critical thinking: What every person needs to survive in a rapidly changing world* (pp. 45-56). Rohnert Park CA: Sonoma State University.

Paul, R. W. (2005). The State of critical thinking today. *New directions for community colleges, 130*, 27-38.

Pellegrino, J. W., & Hilton, M. L. (Eds.) (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* National Academies Press.

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational Leadership*, 46, 22-32.

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research, 42,* 237-249.

Prowse Turner, J. A., & Thompson, V. (2009). The role of training, alternative models, and logical necessity in determining confidence in syllogistic reasoning. *Thinking and Reasoning, 15*, 69-100.

Rachlinski, J. J. (2004). Heuristics, biases, and governance. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell Handbook of Judgment & Decision Making* (pp. 567-584). Malden: Blackwell Publishing.

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

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*, 529–556.

Renkl, A. (2005). The worked-out principle in multi-media learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229-246). Cambridge: Cambridge University Press.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary educational psychology, 23*, 90-108.

Ritchart, R., & Perkins, D. N. (2005). Learning to think: The challenges of teaching thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 775- 802). New York: Cambridge University Press.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*, 1-15.

Roediger, H. L. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest, 14*, 1-3.

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271-286). Cambridge: Cambridge University Press.

Sá, W., Kelley, C., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences, 38*, 1149-1161.

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.

Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. (2006). Neuroeconomics: Cross-currents in resaerch on decision-making. *Trends in Cognitive Sciences, 10*, 108-116.

Schagen, I., & Elliot, K. (2004). *But what does it mean? The use of effect sizes in educational research*. London: National Foundation for Educational Research.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education, 46*, 426–445.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*, 285–296.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34*, 619-632.

Siegel, H. (1980). Critical thinking as an educational ideal. *The Educational Forum, 45*, 7-23.

Siegel, H. (1989). The rationality of science, critical thinking, and science education. *Synthese, 80*, 9-41.

Smith, G. F. (2003). Beyond critical thinking and decision making: Teaching business students how to think. *Journal of Management Education*, 27, 24-51.

Stanovich, K. E. (2009). Distinguishing the reflective algorithmic and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual Processess and beyond* (pp. 55-88). Oxford: Oxford University Press.

Stanovich, K. E. (2011). *Rationality & the Reflective Mind.* New York: Oxford University Press.

Stanovich, K. E., & Stanovich, P. J. (2010). A frame-work for critical thinking, rational thinking and intelligence. In D. D. Preiss & R. J. Sternberg (Eds.), *Educational* *psychology; Perspectives on Learning, Teaching and Human Development* (pp. 195-237). New York: Springer Publishing Company.

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristic and biases. *Advances in Child Development and Behavior, 36*, 251-285.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342-357.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking and Reasoning, 13*, 225-247.

Ten Dam, G., & Volman, M. (2004). Critical thinking as a citizenship competence: Teaching strategies. *Learning and Instruction, 14*, 359-379.

Thompson, V. (2009). Dual processes theories; A metacognitive perspective. In J. Evans, & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). New York: Oxford University Press.

Thompson, V., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63,* 107- 140.

Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional case studies. *The Journal of Higher Education, 73*, pp. 740-763.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16-26.

Wason, P. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*, 63-71.

Contents

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 509-521.

Wenzlaff, R. M., & Bates, D. E. (2000). The relative efficacy of concentration and suppression strategies of mental control. *Personality and Social Psychology Bulletin, 26*, 1200-1212.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*, 930-941.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science, 34*, 776-806.

Wolcott, S. K., Baril, C. P., Cunningham, B. M., Fordham, D. R., & St.Pierre, K. (2002). Critical thought on critical thinking research. *Journal of Accounting Education, 20*, 85-103.

# Samenvatting

Het hoger onderwijs staat voor de uitdagende taak om studenten voor te bereiden op de complexe wereld van de 21ste eeuw, waarin kritisch denken een essentiële vaardigheid is (Pellegrino & Hilton, 2012). Kritisch denken wordt onder meer in verband gebracht met het nemen van weloverwogen beslissingen, leren en transfer (Facione, 1990; Helsdingen, Van Gog, & Van Merriënboer, 2011). Kritisch denken kan worden gedefinieerd als het vermogen om "doelgericht, zelfregulerend te oordelen, resulterend in interpretatie, analyse, evaluatie, het trekken van conclusies, alsook het verklaren van de bewijzen, de concepten, de methode, de criteria en de contextuele overwegingen waarop het oordeel gebaseerd is" (Facione, 1990). Er zijn veel verschillende definities van kritisch denken die een grote mate van overlap vertonen. Maar er zijn ook verschillen en een zo'n verschil, waarover tot op heden discussie is, gaat over de vraag of kritisch denken een domein-specifieke of domein-generieke vaardigheid is (Davies, 2013). De studies in dit proefschrift sluiten aan bij een domein-generiek perspectief waarbij kritisch denken wordt gerelateerd aan het begrip rationaliteit. Een rationeel persoon handelt en laat zich overtuigen op basis van redenen (Siegel, 1989). Het is de taak van het onderwijs om deze rationaliteit te bevorderen door bijvoorbeeld studenten te leren om onbevooroordeeld te redeneren zonder systematische redeneerfouten te maken. Met andere woorden: om te redeneren zonder cognitieve bias. Dit onderwerp heeft tot nu toe nog vrij weinig aandacht gekregen in de 'kritisch denken' literatuur (West, Toplak, & Stanovich, 2008).

Bias in redeneren komt veel voor en is overtuigend aangetoond in bijvoorbeeld de 'heuristics and biases' traditie door onder meer Kahneman en Tversky (1981). Bias kan verklaard worden door twee typen denkprocessen: Type 1 processen verlopen spontaan en vragen weinig capaciteit van het werkgeheugen, in tegenstelling tot Type 2 processen die gekenmerkt worden door weloverwogen denkstappen die een sterk beroep doen op het werkgeheugen (Evans, 2011). Type 1 processen zijn erg efficient voor het uitvoeren van routine taken, maar zetten ook de deur open voor bias, tenzij Type 2 processen in werking treden (Stanovich, 2011). Om Type 2 processen te activeren dient men zowel te beschikken over kennis en strategieën van redeneerregels en –principes, als de denkhouding (dispositie) te hebben om de betreffende regels en principes toe te (willen) passen (Stanovich, 2011). De veronderstelling is dat instructies Type 2 processen kunnen aanmoedigen en daardoor bias kunnen voorkomen, maar de vraag is welke instructiemethoden daarvoor het meest effectief zijn (Larrick, 2004; Stanovich & Stanovich, 2011).

Deze vraag is vooral belangrijk in programma's in het hoger onderwijs, die tot doel hebben om studenten voor te bereiden op het werken in dynamische en complexe professionele omgevingen, zoals in de economische, de juridische of de medische sector. Onjuiste beslissingen als gevolg van bias kunnen in die sectoren zeer ernstige gevolgen hebben in termen van financiële verliezen, emotionele schade, en zelfs mensenlevens (zie bijvoorbeeld: Chapman & Elstein, 2000; De Bondt en Thaler, 2002; Tversky & Kahneman, 1974).

Het doel van de drie empirische studies in dit proefschrift was dan ook om de impact van verschillende instructies te onderzoeken op een essentieel aspect van het kritisch denken, namelijk het vermogen om zonder bias te redeneren. Het eerste doel was om de effecten te onderzoeken van algemene expliciete instructies in kritisch denken (vergeleken met impliciete instructies of geen instructies) gecombineerd met het oefenen in de context van een domein (met en zonder specifieke aanwijzingen om diepgaand redeneren te stimuleren), op de redeneervaardigheden van economie studenten in het hoger beroepsonderwijs. De specifieke aanwijzingen bij het oefenen bestonden uit de vraag om te verklaren hoe men op het antwoord was gekomen ( = self-explanation) of uit het vestigen van de aandacht op relevante informatie in de taak (= activation). Het tweede doel was om de rol van de denkhouding te onderzoeken en om eventuele interacties tussen de denkhouding en de instructies op te sporen. Het derde doel was om de rol van zekerheid en mentale inspanning voor en na de instructies te onderzoeken. De drie experimenten werden uitgevoerd in drie academies binnen de sector Hoger Economisch Onderwijs van een Nederlandse Hogeschool.

Het effect van expliciete instructies werd onderzocht in de studies in Hoofdstuk 2, 3 en 4. De studie in Hoofdstuk 2 liet zien dat impliciete instructie in een bestaande cursus over argumenteren en onderhandelen niet voldoende was om het redeneren zonder bias te verbeteren. Alleen expliciete instructie in kritisch denken gecombineerd met het oefenen van taken in een domeincontext was effectief. Er was geen verbetering op taken die geïnstrueerd maar niet geoefend waren of op taken die geoefend maar niet geïnstrueerd waren. Aanwijzingen tijdens het oefenen ('self-explanation' of 'activation') verbeterden het redeneren niet.

De studie in Hoofdstuk 3 liet opnieuw zien dat een combinatie van expliciete instructie in kritisch denken en het oefenen van taken in een domeincontext tot betere redeneerprestaties leidden vergeleken met redeneerprestaties in een controle groep, ook na een periode van drie weken. Opnieuw lieten studenten, vergeleken met de controlegroep, geen beter resultaat zien op taken die geïnstrueerd maar niet geoefend waren of op taken die wel geoefend maar niet geïnstrueerd waren. Er was echter één uitzondering: studenten die instructies kregen in kritisch denken, plus oefenden op taken, plus tijdens oefenen aanwijzingen kregen om te verklaren hoe ze aan het antwoord waren gekomen ('self-explanation'), scoorden beter op taken die niet geoefend waren dan studenten in alle andere condities. Dit kan duiden op een 'transfer

Contents

effect' van geoefende naar niet geoefende taken, alhoewel dit effect na drie weken verdwenen was. Omdat uit de studies in Hoofdstuk 2 en 3 bleek dat expliciete instructie plus het oefenen de redeneerprestatie verbeterde, rees de vraag of dit kwam door de specifieke invloed van het oefenen in een domeincontext of doordat studenten in die condities meer tijd konden besteden aan het leren.

Dit werd onderzocht in de studie in Hoofdstuk 4. Daarnaast werd in de studie in Hoofdstuk 4 onderzocht in hoeverre het 'transfer' effect (d.w.z. betere prestatie op niet-geoefende taken), dat naar voren kwam in de studie in Hoofdstuk 3, afhankelijk was van het al dan niet vooraf instructie krijgen op de betreffende taken. De analyses in de studie in de studie in Hoofdstuk 4 lieten zien dat oefenen geen beter resultaat opleverde dan een aanvullende diepgaande verwerking van de instructie. Bovendien bleek dat de 'self-explanation' aanwijzing gedurende het oefenen geen effect had op getrainde taken (geïnstrueerde en/of geoefende taken) zoals in Hoofdstuk 2 en 3, noch op niet getrainde taken. Dus als taken niet geïnstrueerd waren leverde de 'self-explanation' aanwijzing geen 'transfer' effect op.  In de studie in Hoofdstuk 4 werd opnieuw aangetoond dat studenten alleen tot betere redeneerprestaties kwamen in de condities waarin ze expliciete instructies in kritisch denken kregen. De bovenstaande resultaten bevestigen en verbreden eerdere onderzoeksbevindingen.

Overeenkomstig met resultaten van onderzoek naar andere aspecten van het kritisch denken (Arum & Roksa, 2011; Bangert-Drowns & Blankert, 1990) laten de bevindingen in dit proefschrift zien dat het redeneren zonder bias zich niet spontaan ontwikkelt als een bijeffect van onderwijs, maar dat expliciete instructies in kritisch denken nodig zijn. Bovendien, in lijn met resultaten van onderzoek naar andere aspecten van kritisch denken (Abrami et al., 2008; Angeli & Valenides, 2009), laten onze bevindingen zien dat voor het verbeteren van redeneren, expliciete algemene instructies gecombineerd met het oefenen in een domeincontext tot betere redeneerprestaties leiden, alhoewel ook het diepgaand verwerken van expliciete instructies in kritisch denken effectief bleek te zijn. Dit strookt ook met inzichten vanuit het duaal-proces model dat er van uit gaat dat Type 1 processen alleen overruled kunnen worden door Type 2 processen als de nodige 'mindware' (kennis en strategieën) aanwezig is. Het geven van de 'self-explanation' aanwijzing tijdens het oefenen, waarvan verwacht werd dat het bij zou dragen aan het juist gebruiken van de aanwezige kennis (Roy & Chi, 2005; Lombrozo, 2006) had geen effect op geoefende taken. Blijkbaar was de instructie met de oefening afdoende om tot betere redeneerprestaties te komen en werkt deze aanwijzing, indien geen instructie vooraf wordt gegeven, niet omdat studenten dan de noodzakelijk informatie missen waarop ze de verklaring kunnen baseren. Zoals ook aangetoond is in eerder onderzoek (Aleven & Koedinger, 2002; Renkl, Stark, Gruber, & Mandl, 1998; Rittle-Johnson, 2006), suggereert de studie in Hoofdstuk 3 dat er een 'transfer' effect van 'self-explanation' is, maar dit dient met voorzichtigheid geïnterpreteerd te worden, aangezien het effect niet aanwezig was op de uitgestelde test na 3 weken en het niet gevonden werd in de studie in Hoofdstuk 2. Het geven van de 'activation' aanwijzingen, waarvan verwacht werd dat het tot betere redeneerprestaties zou leiden (Moutier, Angeard, & Houdé, 2002; Moutier & Houdé, 2003), had geen effect. Blijkbaar was de instructie en het oefenen voldoende om de aandacht te richten op relevante taakaspecten en om automatische reacties te remmen.

De impact van de denkhouding werd onderzocht in de studies in Hoofdstuk 2, 3 en 4. Verwacht werd dat het actief zoeken naar bewijs tegen de eigen opvattingen, plannen, en doelen en het vermogen om bewijs in alle redelijkheid af te wegen (AOT: Actively Open-minded Thinking) zou leiden tot een hogere scores in redeneer prestaties. Dit werd bevestigd in de studies in zowel Hoofdstuk 2, 3 als 4 en is overeenkomstig eerder onderzoek (Macpherson & Stanovich, 2007;Sá, Kelley, Ho, & Stanovich, 2005; Stanovich & West, 1997; West, Toplak, & Stanovich, 2008). Echter, studenten met hogere scores op NFC (Need for Cognition; dat wil zeggen de neiging hebben om door te denken en daaraan plezier te ontlenen) scoorden alleen hoger op redeneerprestaties in de studie in Hoofdstuk 3, en niet in de studies in Hoofdstuk 2 en 4. Mogelijk dat de experimentele setting in de studies in Hoofdstuk 2 en 4 de potentiële effecten van NFC heeft overvleugeld; NFC is sterk gerelateerd aan cognitieve motivatie (Cacioppo, Petty, Feinstein, & Jarvis, 1996). Misschien dat studenten in de studie in Hoofdstuk 3 een dergelijke externe regulering in mindere mate hebben ervaren, zij waren deeltijd studenten in tegenstelling tot de studenten in de studies in Hoofdstuk 2 en 4.

Een nog interessantere vraag, was echter of studenten die hoog scoren op deze denkhoudingaspecten ook meer zouden leren van de instructies dan studenten die laag scoren op deze denkhoudingaspecten. In alle drie de studies (Hoofdstuk 2, 3 als 4) werd er geen interactie met instructies gevonden, oftewel: onder de condities waarin expliciete instructies werden gegeven profiteerden alle studenten evenveel van de instructies die werden aangeboden. Dit lijkt goed nieuws voor docenten, hoewel studenten met lagere scores op disposities misschien meer aandacht nodig hebben om ze uiteindelijk op hetzelfde hoge niveau te krijgen.

Om de potentiële effecten van instructies op de belasting van het werkgeheugen (hierna: cognitieve belasting) te verkennen, werd in de studies in Hoofdstuk 2 en 4 de mentale inspanning van studenten in de verschillende condities onderzocht. Enerzijds zou kennis van strategieën verworven door instructies de cognitieve belasting kunnen verlagen en zouden instructies daarom tot minder mentale inspanning kunnen leiden op de nameting (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Anderzijds zouden instructies er toe kunnen leiden dat studenten zich bewust worden van mogelijke bias en daarom meer mentale inspanning leveren (Evans, 2011). In beide studies werden er

Contents

geen verschillen gevonden in mentale inspanning tussen condities op de redeneertaken die geïnstrueerd, geoefend of getraind waren. Gegeven dat studenten in condities met expliciete instructies in kritisch denken, met oefeningen in de domeincontext of het opnieuw bestuderen van de instructie, een beter resultaat lieten zien terwijl ze dezelfde mate van mentale inspanning rapporteerden, betekent waarschijnlijk dat ze zich evenveel inspanden, maar voor andere denkprocessen. Met andere woorden: de instructie beïnvloedde de efficiëntie van de redeneerprestatie: een beter resultaat werd bereikt met een vergelijkbare inspanning na instructie (Hoffman & Schraw, 2010; Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). Interessant is dat in de studie in Hoofdstuk 4 alleen de studenten die expliciete instructie in kritisch denken kregen (op de helft van de taken) een verhoogde mentale inspanning rapporteerden op niet getrainde (d.w.z.: niet geoefende en niet geïnstrueerde taken) terwijl de prestatie niet verbeterde. Dit kan betekenen dat de instructie de studenten heeft gestimuleerd om niet te reageren op de 'automatische piloot' (Type 2 proces in plaats van Type 1 proces; Evans, 2011), maar dat de kennis ontbrak om het juiste antwoord te geven.

Uit de studie in hoofdstuk 4 kan worden afgeleid dat zekerheid en mentale inspanning negatief gecorreleerd waren; hoe zekerder een student was over het antwoord hoe minder inspanning er door de student werd geleverd. Dit kan worden verklaard door het 'fluency' principe (Thompson, 2009), dat stelt dat taken die een snelle response uitlokken, een gevoel van 'zekerheid' veroorzaken, dat vervolgens een signaal afgeeft dat verdere inspanning niet meer nodig is. Het kan ook verklaard worden door het 'sufficiency' principe (Chen & Chaiken, 1998) dat inhoudt dat men geneigd is om slechts zo lang inspanning te leveren tot men zich zeker voelt. Op de voormeting waren de prestatie op redeneertaken en zekerheid niet gecorreleerd, maar op de nameting was er een positieve relatie tussen de prestatie op redeneertaken en zekerheid. Dit duidt er misschien op dat de instructies de studenten iets voorzichtiger hebben gemaakt want studenten rapporteerden ten opzichte van de voormeting minder zekerheid op niet getrainde taken en evenveel zekerheid op getrainde taken.

Kortom de studies in dit proefschrift laten zien dat expliciete instructies in kritisch denken noodzakelijk zijn maar waarschijnlijk niet voldoende om bias in redeneren te voorkomen. Instructies in kritisch denken waren in de studies van Hoofdstuk 2 en 3 alleen effectief als ze gecombineerd werden met het oefenen in een domein context, hoewel uit de studie in Hoofdstuk 4 blijkt dat het actief verwerken van tekstuele instructies net zo effectief was als het oefenen in een domeincontext. We vonden geen bewijs voor transfer naar taken die geïnstrueerd maar niet geoefend waren (Hoofdstuk 2 en 3) of naar taken die sowieso niet getraind (geïnstrueerd of geoefend) waren. Alleen de 'self-explantion' aanwijzing leek een 'transfer' effect te hebben naar taken die geïnstrueerd maar niet geoefend waren, hoewel dit alleen werd gevonden op de test die onmiddellijk volgde na de interventie en niet op de uitgestelde test en alleen in de studie in Hoofdstuk 3.

De praktische implicaties van dit proefschrift kunnen als volgt samengevat worden: Docenten kunnen en dienen een actieve rol te hebben in het 'onderwijzen' van studenten in het redeneren zonder bias. Kritisch denken ontwikkelt zich niet spontaan en dient ingebed te worden in hoger onderwijs curricula. Dit is niet alleen een verantwoordelijkheid van docenten maar ook van onderwijsontwikkelaars en adviseurs die betrokken zijn bij onderwijsinnovatie en evaluatie. Wat hun taak bemoeilijkt, is dat het niet altijd even duidelijk is wat werkt, en wat niet werkt. Dit proefschrift laat zien dat studenten alleen leren als een combinatie van expliciete instructie wordt gecombineerd met activiteiten die een diepgaande verwerking van de instructie mogelijk maken. Bovendien vereist het geven van instructies over de menselijke neiging om niet ten prooi te vallen aan cognitieve bias, van docenten dat ze zelf die bias en valkuilen begrijpen en kunnen vermijden. Kritisch denken zou dus een belangrijk onderdeel moeten worden van docent professionalisering. Veelbelovend is het gegeven dat in dit proefschrift, de redeneervaardigheden al verbeterden na betrekkelijk eenvoudige en korte interventies, die gemakkelijk te implementeren en te herhalen zijn in het curriculum.

In Hoofdstuk 5 worden ook aanbevelingen gedaan voor toekomstig onderzoek, met name om te verkennen hoe de instructies geïntegreerd kunnen worden in een langlopend curriculum. Bijvoorbeeld is het de vraag of het effectief zou zijn om studenten gedurende een hele cursus of meerdere cursussen aan expliciete instructies en oefenen bloot te stellen. Herhaalde instructie gedurende een cursus of gedurende een heel curriculum kan mogelijk een sterker effect hebben, en kan economie studenten misschien helpen om bias in redeneren te voorkomen als ze uiteindelijk in dynamische en complexe business omgevingen werken. Met het oog op dat laatste, zou het ook interessant zijn om verder te onderzoeken hoe instructie de transfer van het geleerde kan versterken; mogelijk moeten studenten expliciete aanwijzingen krijgen dat ze hetgeen ze geleerd hebben ook op andere taken kunnen toepassen. Verder is het interessant om na te gaan hoe de denkhouding verder versterkt kan worden door instructies. Bijvoorbeeld, zou het met regelmaat instrueren van studenten, ook invloed hebben op hun overtuigingen, cognitieve stijl, doelen, of de waarde die ze aan kennis hechten? Gelet op het belang van kritisch denken zou het ultieme doel van onderwijs moeten zijn om studenten af te leveren die kritisch kunnen denken in het onderwijs en in de professionele context zonder dat ze daar aanwijzingen voor krijgen.

Contents

# Presentations in the Context of this Dissertation

Heijltjes, A., van Gog, T., & Paas, F., (2010, April). *Kritisch leren denken van economie studenten*. Paper presented at the Jaarcongres HBO-raad, Amersfoort, the Netherlands.

Heijltjes, A., van Gog, T., & Paas, F., (2010, June). *How to instruct economics students to think critically? Efforts to overrule the autonomous mind.* Paper presented at Edineb, London, United Kingdom.

Heijltjes, A., van Gog, T., & Paas, F., (2011, September). *Effects of instructions on economics students' critical thinking*. Paper presented at the meeting of the European Association for Research on Learning and Instruction, Exeter, United Kingdom.

Heijltjes, A., van Gog, T., & Paas, F., (2012, April). *Improving economics and business students' critical thinking: Empirical support for explicit instructions*. Paper presented at the 5th Cognitive Load Theory conference, Tallahassee, Florida, USA.

Heijltjes, A., van Gog, T., & Paas, F., (2012, April). *Improving economics and business students' critical thinking: Empirical support for explicit instructions.* Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Heijltjes, A., van Gog, T., Leppink, J., & Paas, F., (2013, August). *Unraveling the effects of critical thinking and practice on students' reasoning*. Paper presented at the meeting of the European Association for Research on Learning and Instruction, Munich, Germany.

# Dankwoord

Voor de steun en betrokkenheid bij de totstandkoming van dit proefschrift wil ik graag een aantal mensen hartelijk danken.

Allereerst mijn promotoren Fred Paas en Tamara van Gog. Fred, dank voor je deskundige begeleiding, je no-nonsense adviezen, je oog voor details, de snelle reacties op mijn vragen en schrijfsels en je bemoedigende woorden als het even tegen zat. Tamara, dank voor je kritische, opbouwende respons, de rake 'voorbeelden', je aanwijzingen bij het schrijven van artikelen, je steun bij presentaties en het vertrouwen dat je in mij stelde. Door jouw relativerende opmerkingen viel er ook heel wat te lachen. Fred en Tamara, na vrijwel elk gesprek op de Erasmus Universiteit vertrok ik met nieuwe inzichten en ideeën die zowel het proefschrift als mijn werk in het Leer- en innovatiecentrum van Avans Hogeschool ten goede kwamen.

Jimmie Leppink, dank voor je waardevolle ondersteuning bij de statistische analyses, je duidelijke en geduldige uitleg en je co-auteurschap.

De leden van de promotiecommissie, Jos Beishuizen, Remy Rikers en Jeroen Van Merriënboer, dank ik voor het lezen en beoordelen van mijn proefschrift. Ook de overige leden van de promotiecommissie, Wolter Paans en Sofie Loyens, dank ik voor de bereidheid om met mij in discussie te gaan over de inhoud van dit proefschrift.

Veel collega's van Avans Hogeschool ben ik erkentelijk voor hun bijdrage: het College van Bestuur voor de promotiefaciliteiten; de academiedirecteuren Jan van Wijk, Noortje de Bruin en René Schenk voor de gelegenheid om binnen hun academies onderzoek te doen. Jan, het thema van dit proefschrift is mede ingegeven door jouw scherpe observaties bij het lezen van afstudeerwerkstukken; docenten en medewerkers van academies en ondersteunende diensten voor de hulp bij de afstemming van cases op het economisch domein, bij de uitvoering van de experimenten in opleidingen en bij het digitaal verzamelen en verwerken van gegevens. In het bijzonder dank ik Jan Baccarne, Menno Broere, Daan Poldermans, Ad van Oosterhout, Angelique Botman, Leo Broekhuis, Johan de Vries, Johan Baardemans, Wim Janssen, Nicole Pas, Kim Anrar, Hugo Kerckhaert en Bert Leene; Erik Blokland en Jet Schmutzer dank ik voor de inspirerende en humorvolle discussies over de link tussen onderzoek en onderwijs.

Mijn collega's van het Leer- en innovatiecentrum hebben op velerlei manieren enthousiast meegewerkt. Ik dank het management, in het bijzonder Rien, Ellen, Joke (Snippe), Maya, Jan (Snijders) en Esther, die het mij mogelijk maakten om werk en promotieonderzoek te combineren. Adri, Asha, Jan (Maas), Karin, Marijke, Emy, Eky, Han, Kees, Marga, Nanny, Arold, Monique, Marian, Niek, Henk, Milou, Floor en Marjolein, bedankt voor alle praktische hulp bij de experimenten. Henk, ook dank voor je input bij de verwerking van data. Arold, ik heb waardering voor je bereidheid om op elk tijdstip materialen in te voeren. Jan (Maas) en Etienne, bedankt voor jullie hulp bij statistische technieken. Milou en Noortje, wat is het fijn om met zulke jonge enthousiaste collega's te mogen werken. Brian, Désiré en Viola, dank voor jullie praktische adviezen en de morele steun. Speciaal dank aan Han: jouw motiverende woorden waren destijds een belangrijke aanjager voor het starten van mijn promotietraject. Fijn dat je mijn paranimf wil zijn. Ook de andere collega's van het Leer- en innovatiecentrum dank ik voor alle belangstelling!

Een woord van dank aan de studenten van Avans Hogeschool die geparticipeerd hebben in het onderzoek. Ook Folkert, bedankt voor je hulp als student-assistent.

Paula en Pieter, dank voor de pakkende vormgeving van de video-instructie: wat zijn jullie daar goed in! Ook bedankt voor jullie belangstelling en vriendschap. Mijn lieve vriendinnen Ellen, Bets, Tineke en Ans, bedankt voor de gezellige weekenden die een welkome afwisseling waren in het straffe tempo van werken en studeren. We maken die uitstapjes al 100 jaar en ik hoop dat we er nog 100 jaar mee doorgaan!

Ik dank mijn ouders en verdere familie voor alle warme support. Lieve Ma, jij bent het beste voorbeeld dat je nooit te oud bent om te leren. Wat jammer dat pa het eindpunt van dit promotietraject niet meer kan meemaken.

Lieve Martijn en Martine, Janneke en Maarten, Riemer en Héloïse, jullie zijn mij dierbaar. Martijn, dank voor het prachtige ontwerp van de cover. Janneke, ik geniet van je brede belangstelling en wat fantastisch dat je mijn paranimf bent. Riemer, dank voor je interesse en je spontane reacties als ik bij jou weer eens een idee toetste om te kijken hoe het zou 'vallen' bij studenten (Nee mam, dat kan echt niet, dan kun je beter…).

Tenslotte Wim, dank voor al je begrip, je geduld, je humor, de welkome wandel- en fietsvakanties en bovenal je liefde. Zonder jou was dit proefschrift er niet gekomen. Dank voor alles!

Contents

# Curriculum Vitae

Anita Heijltjes was born in Zevenaar, the Netherlands, on July 18th 1955. She studied Nursing and obtained a first level teaching qualification in Nursing, and a master degree in Health Sciences at Maastricht University. Since 1982, she has worked in the field of higher professional education, initially at Hogeschool Zuyd, and since 1983 at Avans Hogeschool (both Universities of Applied Sciences). She has held various positions; as lecturer at the Bachelor of Nursing, as project leader involved in restructuring regional health care educational programmes (West Brabant) and as head of Department of Higher Education in Nursing. Since 2004, she has worked at the Learning and Innovation Centre of Avans Hogeschool as a senior educational policy advisor to support management and faculty in developing educational innovations. She was involved as project leader of the curriculum revision of the teacher-training department (Pabo). In January 2013, she initiated a research group ‘Brain & Learning’. In this research group, her focus is on translating insights about the learning brain and the cognitive architecture of memory into everyday educational practice. In 2008, she started her (part-time) PhD study at Erasmus University Rotterdam regarding the learning of economics students’ critical thinking skills. The main themes of her career have been creating optimal conditions for learning and connecting science with practice.