Instructional Design Principles to Reduce Confirmation Bias

Suzan van Brussel

layout and print: proefschrift-aio.nl

Instructional Design Principles to Reduce Confirmation Bias

Ontwerpprincipes voor het onderwijs om de voorkeur voor bevestiging te verminderen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
14 april 2022 om 13 uur

door

Suzan van Brussel
geboren te Beek en Donk.

Erasmus University Rotterdam

Promotiecommissie:

Promotor:	prof. dr. F. Paas
Overige leden:	prof. dr. L. R. Arends prof. dr. T. A. J. M. van Gog prof. dr. J. W. F. van Tartwijk
Copromotoren:	dr. P. P. J. L. Verkoeijen dr. M. C. L. Timmermans

Table of Contents

Chapter 1
General introduction

We live in an era with a tremendous exposure to information. Much of the information that will pop up on our screens, echoes our own point of view and we hardly perform reflective thoughts on what we read and see (Quatrociocchi et al., 2016). Moreover, algorithms are used to present the information that confirms our points and omits alternatives (Vydiswaran et al., 2015; Zimmer et al., 2019). This may result in one-sided views, thereby hampering unbiased reasoning (Zimmer et al., 2019). Flaws in human reasoning are, however, of all times. To ease our reasoning processes, we rely on heuristics (Tversky & Kahneman, 1974). Heuristics are mental shortcuts based on previous experiences and existing knowledge, that allow for split-second decisions and yield accurate judgments in routine situations. Heuristics can be handy mnemonics that guide us quickly to possible solutions and save time and mental effort when not all the necessary information is available. Nevertheless, heuristics can also yield cognitive biases. Before explaining the concepts of heuristics and biases in more detail, let’s consider the following non-routine example in which one has to decide which cards one needs to turn in order to test a rule:

If a card has a vowel on one side, then it has an even number on the other side.
Which card(s) do you need to turn over in order to determine whether the rule is true or false?

A	**D**	**4**	**7**

The type of task used above is named Wason’s four-card selection task (Wason, 1968). Wason found that people in general select the card(s) that confirm(s) the rule and seldomly select a card that might disconfirm the rule i.e., the heuristic response that most people choose to turn over the vowel A, and the even number 4. The correct response, though, is to turn over A and 7. The rule was “***If*** a card has a vowel on one side, ***then*** it has an even number on the other side”. Only a card with both a vowel on one side **and** an odd number on the other side can invalidate the rule. Finding an odd number on the other side of the A or finding a vowel behind the 7 would reveal the statement to be false. The cards showing B and 4 are incorrect selections, because whatever is on their other sides is consistent with the rule. In short, one can determine the claim to be false by finding either the card showing the A or the card showing the 7 to be inconsistent with it, or one can determine the claim to be true by finding both of these cards to be consistent with it (Nickerson, 1998, p. 184). Therefore, solving the task requires replacement of the heuristic response with a response based on the rules of formal logic, namely an implication rule of the conditional “If P happens, then Q happens”. The basic finding of experiments with this type of task supports the hypothesis that people find the modus tollens (i.e., reject) argument to be less natural than the modus ponens (i.e., confirm) form (Evans & Stanovich, 2013; Nickerson, 1998; See for an explanation of modus tollens and modus ponens: Torringa, 2019). In the example, P is the card with the vowel A (modus ponens) and the card with the consonant D is not-P (denying the antecedent). The card with the even number 4, is Q (affirming the consequent), and the card with the uneven number 7, is not-Q (modus tollens). The correct answer to these kinds of tasks is therefore "P and not-Q" since only the combination of P and not-Q can falsify an implication rule. This logical rule goes against people's general preference for confirmation (P and Q). Hence, it might not come as a surprise that people in general perform poorly on this type of task (Nickerson, 1998; Ragni et al., 2018; Wason, 1968).

Heuristics and biases

People's preference for confirming the rule in for instance the task example above, can be explained by heuristics, which result in biases (De Neys, 2021; Evans, 2003; Evans & Stanovich, 2013; Stanovich, 2011). If one tries to analyse every aspect of every situation or decision that one is confronted with, daily life would be an obstacle course. That is why heuristics are very useful. Tversky and Kahneman (1974) indicated three heuristics: availability, representativeness, and anchoring-and-adjustment. An example of the first heuristic is when people evaluate the probability of an event according to the ease with which examples come to mind, i.e., the availability heuristic (Kahneman, 2011; Tversky & Kahneman, 1974). As an illustration, people estimate the risk of an airplane crash higher than the probability of a car crash. It seems as though airplane crashes are happening more frequently than they actually are because when an airplane crashes, it is widely reported throughout the media leading to stronger memories compared to car crashes. However, the likelihood of another crash has not changed, rather the awareness of airplane accidents has increased. The frequency with which these events occur is overestimated and the use of this heuristic therefore results in biased reasoning. In contrast, the representative heuristic is present when people use existing memories to identify associated characteristics of an object or a person. An example is the "Linda" conjunction fallacy task (Tversky & Kahneman, 1983) in which people receive the following description: 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. Afterwards, people need to decide whether it is more probable that Linda is a bank teller, or a bank teller and a feminist movement activist. Most people choose the second option because the "conjunction" represents the description more than the first option. This option is, however, mathematically less likely. In short, the probability of two events occurring together can never be larger than the probability of any of these events occurring alone. The third heuristic is the anchoring-and-adjustment heuristic (Tversky & Kahneman, 1974). For example, when people think the outcome of 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 is higher than the outcome of 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8, it is because people focus on the initial number (the anchor) and estimate the outcome towards the anchor (adjustment).

We use heuristics as a cognitive miser by reducing the cognitive effort that is required to make decision and judgments, when faced with a complex decision to decide earlier and faster, and as rules of thumb because heuristics are usually correct (Stanovich, 2011; Tversky & Kahneman, 1974). Despite the advantages in decision time and effort, a heuristic can also open the door to systematic thinking errors, that is to say cognitive biases. Cognitive biases are the result of applying heuristics during reasoning, judgment and decision making, and they appear when we draw inferences or adopt beliefs where the evidence for doing so in a rationally sound manner is either insufficient or absent (Haselton et al., 2015).

The appearance of cognitive biases can be explained by dual process theories (De Neys, 2021; Evans, 2003; Evans & Stanovich, 2013; Stanovich, 2011). The dual processing theory by Evans (2003), gives insight in the underlying cognitive thinking processes of biased thinking based on two types of reasoning processes. Dual process theories propose that humans have two kinds of information processing at their disposal: Type 1 and Type 2 thinking, sometimes referred to as System 1 and System 2 thinking (Kahneman, 2011; Stanovich & West, 2008). Type 1 thinking processes are automatic, rely on heuristics, require little effort and are efficient in many routine situations (Evans, 2003). Type 2 thinking processes are more deliberate, reflective thinking systems that refer to rationality and sequential thinking. Excessive reliance on Type 1 thinking processes can overrule Type 2 thinking processes, preventing reflection and leading to unexamined decisions. For example, solving 2 + 2 or crossing a nearby street, appeals to Type 1 thinking. One cannot turn off Type 1 thinking: One of the tasks of Type 2 thinking is to suppress Type 1 thinking when deliberate judgments or decisions are required (Kahneman, 2011). Examples of situations in which Type 2 thinking is required are for example solving 173 x 245, driving a car through Paris during rush hour, compiling the secondary school advice in group 8 of primary school, or (and much less at stake) solving a Wason's four-card selection task. In non-routine situations when something is at stake, and when information should be critically assessed, people should use reflective Type 2 thinking processes instead. Evans' theory (2003), indicates that thinking errors occur because of rapid, heuristic-based Type 1 reasoning instead of applying the analytic, reflective Type 2 reasoning in high stake decision making and judgment situations. These errors occur in situations in which people either don't realize that a Type 2 response is needed, or they don't apply a Type 2 response because it takes too much effort, or they don't have the right knowledge and skills to generate an appropriate Type 2 response.

The Wason four-card selection task example mentioned above, provides important evidence for dual processing in reasoning because performance on this abstract task is thought to be strongly influenced by a Type 1 heuristic, namely the tendency to focus attention on evidence containing the letters and numbers mentioned in the rule, regardless of the logical rules (Evans, 2003). Retrospectively, you yourself might have realised that you applied the heuristic-based Type 1 thinking processes to solve the aforementioned example of this task, leading to a cognitive bias, namely the preference for confirmation. This cognitive bias is known as the "confirmation bias". This bias refers to the tendency to selectively attend to, remember, and interpret information that is consistent with held beliefs or expectations, or to avoid information that is inconsistent with one's own preferences (Cavojova et al., 2018; Nickerson, 1998; Schwind et al., 2012; Stanovich et al., 2016; Sternberg & Halpern, 2020). Wason's four-card selection tasks are frequently used to measure the confirmation bias (e.g., Evans et al., 1993; Heijltjes, Van Gog, Leppink, et al., 2014; Sellier et al., 2019; Stanovich & West, 2008; Van Peppen et al., 2018; Wason, 1968). We will discuss the confirmation bias in more detail in the next section.

The confirmation bias

Philosopher Francis Bacon (1620 / 2000), explained already in the 17th century that the confirmation bias is an important determinant of our thought and behaviour:

> *"The human understanding when it has once adopted an opinion (either as being the received opinion or as being agreeable to itself) draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate."*
>
> *(Bacon et al., 1620 / 2000, XLVI, p.12)*

The confirmation bias is based on cognitive processes that interfere with unbiased reasoning. People have the tendency to give more attention and place more value on information that confirms their ideas, beliefs, or hypotheses, and place less value and pay less attention to information that contradicts those ideas (Nickerson, 1998). Therefore, they seek confirmatory evidence when evaluating competing hypotheses, they misperceive evidence to support their beliefs, and they place more weight on confirming evidence compared to disconfirming evidence when they update their beliefs (Charness & Dave, 2017; Nickerson, 1998).

In his article, Nickerson (1998) addresses the confirmation bias extensively. He appoints examples of the confirmation bias in real-life situations, e.g., when diagnosing in medicine, reasoning in science and during judicial reasoning (Nickerson, 1998, pp. 192-194). People fall prey to the confirmation bias on the Internet as well. For example, Quatrociocchi et al. (2016), showed that Facebook users most of the time interact with like-minded people in echo chambers, and have a tendency to search for information that supports their preference on a subject, and to reject disproving information. Even worse, the confirmation bias is intensified by the "filter bubble", which means that the algorithm of a search engine only displays information that one is likely to agree with, and blocks contradictory viewpoints and, hence, possibilities for perspective taking (Pariser, 2012).

Participants in this dissertation's experiments were predominantly student teachers from Dutch primary teacher education institutions. In the context of primary education, the confirmation bias might have far reaching consequences. First impressions about e.g., children's socio-economic background or intelligence, that may hold stereotypes and implicit and partly erroneous beliefs, resulting in biased expectations influence teachers' behaviour towards pupils (De Boer et al., 2018; Kelley, 1950; Timmermans et al., 2018). This in turn might lead to a self-fulfilling prophecy in which children's learning and motivation are negatively affected because they become aligned with the teacher's erroneous prior beliefs or stereotypes (Darley & Gross, 1983; Foster et al., 1976; Rosenthal & Jacobson, 1968; Timmermans et al., 2015; Van den Bergh et al., 2010). Ultimately, the teacher's confirmation bias might therefore differentially affect the pupil's academic progress (e.g., Darley & Gross, 1983; De Boer et al., 2010; Foster et al., 1976; Rosenthal & Jacobson, 1968; Timmermans et al., 2015; Van den Bergh et al., 2010; Vanlommel et al., 2018, 2020). Accordingly, when teachers only attend to, or prioritize classroom information that is consistent with their existing beliefs and expectations, their confirmation bias might hamper good judgment and decision making. Their confirmation bias may also result in one-sidedness in courses where perspective taking and open-mindedness are important elements, such as lessons regarding social topics that are commonly addressed in citizenship education such as elections, diversity, or sexual orientation (e.g., Abrami et al., 2015; Nickerson, 1998; Schwind et al., 2012). Hence, it is important for student teachers to learn what the confirmation bias entails, to be aware of this bias, and to know how to reduce it, so that they can effectively deal with confirmation bias as a professional. However, the confirmation bias cannot be eliminated entirely, but people can learn to mitigate and reduce this bias by training and education in critical thinking (Evans, 1996; Kahneman, 2011; Lord et al., 1984; Lord et al., 1979; Nickerson, 1998).

Critical thinking

People have to learn and make use of critical thinking (hereafter, referred to as CT) to learn to reason unbiasedly, hence, to mitigate their confirmation bias. As a result, decision making and judging improve. In recent years, there has been an increasing interest among researchers and teachers in the educational goal of enhancing students' CT (Janssen, Mainhard, et al., 2019; Pellegrino & Hilton, 2012; Ten Dam & Volman, 2004; Yang & Chou, 2008).

Lipman (1988, p. 10), argues that CT is "skilful, responsible thinking that facilitates good judgment because it relies upon criteria, is self-correcting and is sensitive to context". Ennis (1989) assumes that thinking critically is reasonable reflective thinking focused on deciding what to believe or do. In 1990, the concept of CT has been elaborated in great detail by the APA Delphi Panel of 46 philosophical, psychological and educational experts. According to the panel, CT consists of knowledge, skills, and dispositions. The panel composed a generally accepted broad definition, according to which CT is characterized as purposeful, self-regulatory judgment, which results in interpretation, analysis, evaluation, and inference, or contextual considerations upon which that judgment is based (Facione, 1990). More recently, Huang et al. (2016, p. 237), have defined CT as "the ability to apply higher cognitive skills (e.g. analysis, synthesis, self-reflection, perspective taking) and/or the disposition to be deliberate about thinking (being open-minded or intellectually honest), which both lead to actions that are logical and appropriate".

Perspective taking is an essential CT skill and open-mindedness is an essential CT disposition to mitigate the confirmation bias. Perspective taking is defined as "the cognitive capacity to consider the world from another individual's viewpoint" (Galinsky et al., 2008, p. 378). Open-mindedness is defined as one's willingness and ability to consider opposing experiences, beliefs, values, and perspectives and give them a serious, impartial consideration by setting aside one's commitment towards one's own experiences, beliefs, values and perspectives (Baehr, 2011; Kwong, 2016). Like CT in general, students do not apply this higher-order thinking skills spontaneously, nor is it a by-product of education (Behar-Horenstein & Niu, 2011). Instead, it requires explicit instruction and practice with the aim of bridging knowledge gaps, being motivated to reason in an unbiased manner, and identifying situations in which to apply it (Abrami et al., 2015; Abrami et al., 2008; Bangert-Drowns & Bankert, 1990; Heijltjes, Van Gog, & Paas, 2014; Janssen, Mainhard, et al., 2019; Marin & Halpern, 2011; Markovits & Brunet, 2012; Mehta & Al-Mahrooqi, 2015; Niu et al., 2013; Soll et al., 2015; Van Peppen, 2020).

Critical thinking to reduce the confirmation bias

Based on their meta-analysis on strategies to teach CT, Abrami et al. (2015, p. 29) conclude that a "magic recipe" for teaching CT lacks, but that dialogue, authentic instruction and mentoring are promising ingredients to learn to successfully master CT. By taking perspective and being open-minded, one considers alternative explanations, which lead to a more balanced and objective evaluation of the evidence that is needed for decision making or judgment (Hirt & Markman, 1995).

Studies have shown that ways to reduce or mitigate the confirmation bias are e.g., prompting students to take another perspective (Snyder & Swann, 1978), generating alternatives to decision objectives (Soll et al., 2015), taking time to consider reasons why one might be wrong or fail (Koriat et al., 1980), playing a bias-mitigating video game in which participants are confronted with decisions and receive feedback on their degree of bias they exhibited (Morewedge et al., 2015), or the thinking strategy "consider-the-opposite" (Hirt & Markman, 1995; Lord et al., 1984; Lord et al., 1979). The latter is a thinking strategy to learn how to mitigate the confirmation bias by generating counter explanations, i.e., to take opposing information into account when making decisions. Considering opposites works because this strategy encourages learners to direct their attention to alternative evidence that otherwise would not be considered (Danielson & Sinatra, 2017; Larrick, 2004). It is known that presenting opposing information prompts learners to reconsider their current position by taking perspective. Lord et al. (1984), named this strategy "consider-the-opposite", and the central question of this strategy presumably induces reflective, Type 2 thinking processes pertaining to the question; "What are some reasons that my initial judgment might be wrong?" (Larrick, 2004, p. 323). Lord et al. (1984), based their experiment on the assumption that biased assimilation of new evidence occurs when alternative and opposite possibilities are neglected. To test whether drawing explicit attention to opposite information would lead to less confirmation bias, they presented participants from the United States who either strongly supported or strongly opposed capital punishment, two summaries of (bogus) research articles, which were either pro or contra the death penalty. In the articles, a comparison of states with and without capital punishment and murder rates before and after the introduction of the death penalty was provided. Participants were asked to read the two summaries of the articles. After reading, the participants indicated to what extent their current attitudes and beliefs towards the death penalty had changed. After this, they read a more detailed article regarding the articles, and they evaluated how well the studies in the articles were conducted, and how convincing each study seemed as evidence on the issue. In a control condition, participants were only asked to be as objective and unbiased as possible in evaluating the studies. In the experimental condition, participants were asked to "consider-the-opposite" by deciding whether they would have made the same high or low evaluation had exactly the same study resulted in the opposing conclusion. By bringing this under the attention of the participants in the experimental condition, Lord and colleagues reminded them that a different experimental design might have brought different cognitions in mind. For participants who favoured the death penalty, the "consider-the-opposite" instruction resulted in a less extreme belief and attitude change. The reverse was true for the participants in the control condition, they showed no belief and attitude change: They kept having a blind spot for the alternative information. Therefore, it was concluded that biases in social judgment can be corrected by explicitly prompting a strategy such as "consider-the-opposite".

Teaching and learning to reduce the confirmation bias

Various studies have provided important theoretical insights into the mechanisms underlying the confirmation bias and ways to reduce it through "consider-the-opposite" (Hirt & Markman, 1995; Lord et al., 1984; Lord et al., 1979). However, these studies were not designed to gain more knowledge for educational practice, i.e., what instructional design is most effective to support mitigating one's confirmation bias. Using "consider-the-opposite" as part of an instruction in educational practice might therefore be promising for learning to mitigate the confirmation bias. To achieve this educational aim, among other things, an effective instruction is required. A number of issues should be addressed in this instruction: Knowledge about what the confirmation bias is, how this bias manifests in work and life, and how to refrain from it. In addition, practising with confirmation bias tasks should be included in the instruction. There are, however, still all kinds of open questions on how to design an instruction in such a way that the confirmation bias is reduced. The overall aim of this dissertation, therefore, was to acquire more knowledge on effective instructional strategies to mitigate the confirmation bias. Below it is explained which research questions were specifically addressed to promote learning to reduce the confirmation bias.

The dissertation consists of four empirical studies on the effect of three important aspects of instructional strategies to reduce the confirmation bias: The domain-specificity of the learning context, the type of feedback provided after practice and the type of cognitive strategies that learners use to process the information provided in the instruction. The confirmation bias was operationalized as performance on hypothesis testing tasks, for example Wason's four-card selection tasks (Chapter 2, 3, and 4), or the quality of an open-minded lesson plan for a citizenship educational lesson in primary school about a social topic (Chapter 5).

An area of discussion within CT research (accordingly within confirmation bias research), is the extent to which CT must be taught within a domain-specific context (e.g., Abrami et al., 2008; Ennis, 1989; Facione, 1990; McPeck, 1990; Moore, 2004; Tiruneh, Gu, et al., 2018). Researchers argue that thinking is always about something and that foreknowledge on a subject is necessary to do so, and therefore a domain-specific instruction is conditionally to learn to think critically (McPeck, 1990; Nickerson, 1988; Willingham, 2007). According to Abrami et al. (2015), an effective CT instruction is an authentic instruction that uses applied problems presented in a real-world context that are genuine to the learner (i.e., for student teachers, within the specific domain of education). Learning in a domain-specific context might stimulate topic interest and subsequent performance: If the topic at hand sparks interest, it is a predictor of students' motivation and performance (Wijnia et al., 2014). Hence, the degree of domain-specificity might positively affect student teachers' performance on confirmation bias tasks, if the instruction and practice are framed within an educational context. It follows then, that the first research question of this dissertation addresses what the effect is of a domain-specific "consider-the-opposite" instruction and practice in an educational context on reducing the confirmation bias compared to a domain-general instruction and practice.

Another instructional strategy is providing feedback on tasks. Feedback is a powerful learning tool because it contributes to better performance, understanding and application of the learning material (Hattie, 2012; Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2005). Feedback can be defined as the information that is provided by an agent (e.g., the teacher) to a student about his/her performance or understanding with the purpose of promoting learning and achievement (Hattie & Timperley, 2007). There are different types of feedback with different learning effects. For instance, correct answer feedback implies that the learner is provided with the correct answer after completing the task (Butler et al., 2013). This is feedback at its most basic level. Bangert-Drowns et al., (1991) have shown that the effect of correct answer feedback is, at its best, very small compared to no feedback. If learners receive additional information in the feedback message, then feedback becomes more effective for learning (Butler et al., 2013). Specifically, elaborative feedback in the form of a worked example, produces larger effects on learning. Worked examples can be regarded as a form of elaborative feedback because a step-by-step explanation and demonstration of a strategy to arrive at the correct solutions is presented (Hattie & Timperley, 2007; Shute, 2008; Sweller et al., 1998; Van der Kleij et al., 2012). This learning effect is called the worked example feedback and can be explained from the Cognitive Load Theory (Paas, Renkl, et al., 2003; Sweller, 1988; 1998, 2019). Studying worked examples during instruction leads to better learning with less investment of time and mental effort compared to solving conventional problems, i.e., goal-specific problems for which learners have to find the solution themselves (Cooper & Sweller, 1987). Worked examples focus on the processes that lead to the correct answer, which facilitates learners' construction of cognitive schemas about the approach to solve problems. This process can guide learning during subsequent practice and problem solving (Cooper & Sweller, 1987; Paas, 1992; Sweller & Cooper, 1985; Sweller et al., 1998, 2019; Van Gog & Rummel, 2010; Ward & Sweller, 1990). Despite its large effect on student achievement, the role of feedback in an instructional design that aims at reducing the confirmation bias in student teachers, has not yet been investigated. Accordingly, in the second study, the central question was whether providing elaborative feedback on practice tasks in the form of worked examples after a "consider-the-opposite" instruction, resulted in better performance compared to providing correct answer feedback, or no feedback.

Studies on cognitive bias reduction in general have shown that a debiasing effect occurs when providing participants with (1) explicit instructions concerning knowledge and the importance of unbiased reasoning with worked examples, (2) the possibility of practising with relevant (domain-specific) unbiased reasoning tasks and (3) feedback on the practice tasks (Adame, 2016; Dunbar et al., 2014; Heijltjes, Van Gog, & Paas, 2014; Mussweiler et al., 2000; Sellier et al., 2019; Van Peppen, 2020). In the "consider-the-opposite" studies by e.g. Jonas et al. (2001), and Adame (2016), participants were left to their own devices with respect to how they processed the information provided in the instruction, i.e., the knowledge and importance of unbiased reasoning. Nevertheless, learners who are given this freedom of choice, mostly rely on suboptimal learning strategies that lead to superficial learning, such as re-reading, summarizing, or highlighting (Blasiman et al., 2017; Dirkx et al., 2019; Dunlosky et al., 2013; Karpicke et al., 2009; Renkl, 2002). According to Fiorella and Mayer (2016), an important factor that influences learning is how learners process the information that is presented in educational materials and instruction, i.e., how learners try to make sense of the information and foster learning themselves. The most effective learning strategies are generative learning strategies, which involve the integration of new information with existing knowledge. These strategies lead to a learner's effort investment in learning to integrate new information with existing knowledge (Fiorella & Mayer, 2016). According to Fiorella and Mayer (2016), the generative learning theory has its origins in Bartlett's view (1932), that learning is a constructive act in which people make an effort to give meaning by integrating new experiences with their existing knowledge structures or schemas. Generative learning activities, therefore, imply "the active construction of meaning from to-be-learned information by mentally reorganizing it and integrating it with one's existing knowledge" (Fiorella & Mayer, 2016, p. 717). Activities are constructive when by undertaking them, learners produce some sort of additional output that might contain ideas that go beyond the provided information from the instruction (Chi, 2009). For example, learners generate statements, ideas or utterances, which are used for self-explanation. Self-explanation is a learning strategy in which learners generate explanations of the learning material themselves and the explanations are directed to themselves with the aim of making sense of new information (Chi et al., 1989; Dunlosky et al., 2013; Rittle-Johnson et al., 2017). Students learn more when they are prompted to explain new information because this promotes comprehension through the integration of new and prior knowledge, and because it makes learners more aware of their reasoning towards the solution, and of still existing knowledge gaps that need to be repaired (Chi et al., 1989; Dunlosky et al., 2013; Fiorella & Mayer, 2016; Rittle-Johnson et al., 2017).

Learning by teaching on video is a variant of a self-explaining learning activity in which learners study the learning material and explain the content to (fictitious) others on video (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019; Jacob et al., 2020; Kobayashi, 2019). Explaining to others leads to deeper learning compared to more superficial learning strategies (Dunlosky et al., 2013). The learner benefits from explaining to others because he/she has to select the relevant information to include in the explanation, organizing it in a way that it can be understood by others, and by elaborating on the material by incorporating one's existing knowledge, which leads to meaningful knowledge structures (Duran, 2017; Fiorella & Mayer, 2016). Moreover, the beneficial learning effects of teaching on video might be explained by the social presence hypothesis: Feelings of social presence are evoked when one is aware of the (fictitious) presence of others in a(n) (virtual) environment (Biocca & Harms, 2002, October; Gunawardena, 1995; Hoogerheide, Renkl, et al., 2019; Lowenthal, 2012; Slater et al., 1999). In turn, feelings of social presence might evoke arousal, which may then result in better learning and transfer compared to e.g., re-study. If one is stimulated to generate accurate and complete explanations to ensure that one's audience understands the content, then this leads to a more effective use of the working memory capacity which consecutively supports the construction of meaningful knowledge in the long-term memory (Sweller et al., 1998; Thayer, 1967). However, there are studies that did not find a relationship between arousal and task performance, and therefore, it remains unknown whether and why teaching on video might be more effective for learning compared to preparing to teach or self-explanation (Hoogerheide, Renkl, et al., 2019; Renkl, 1995). In sum, although there are still many open questions about the beneficial mechanisms underlying the teaching on video strategy, recent studies have shown promising results compared to only preparing to teach (i.e., teaching expectancy), or re-study, especially for learners with low prior knowledge (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2016; Hoogerheide et al., 2014b; Hoogerheide, Renkl, et al., 2019). However, the effects on learning by teaching on video have not yet been investigated in the context of learning to reduce the confirmation bias through "consider-the-opposite". The aim of the third and fourth experiment, therefore, was to examine the role of learning by teaching on video to process an instruction based on "consider-the-opposite". Learning by teaching on video was pitted against only preparing to teach, and re-study to investigate the effect of these strategies regarding learning to reduce the confirmation bias as measured with confirmation bias tasks (study 3), and the effect of these strategies on preparing open-minded citizenship lessons on a social topic (study 4).

In the next section, the hypotheses we draw from the literature, and the experiments that were conducted to test the hypotheses, are presented.

Overview of the dissertation

The dissertation consists of four empirical studies on the effect of different instructional designs to reduce the confirmation bias. The independent variables were learning context (**Chapter 2**), feedback (**Chapter 3**), and the generative learning strategy "teaching on video" (**Chapter 4 and 5**), and its effects on participants' confirmation bias.

Learning context is examined in **Chapter 2**. The main research question was whether a "consider-the-opposite" instruction and practice in a domain-specific context would lead to better performance on a confirmation bias-posttest compared to instruction and practice in a domain-general context. Student teachers were randomly assigned to the domain-specific Educational Context (EC) or the domain-general General Context (GC) condition. We hypothesized that instruction and practice in a domain-specific, educational context would reduce student teachers' confirmation bias more than instruction and practice in a general context because an authentic learning context has been shown to spark interest more; EC > GC. The experiment used a 2 (Context: EC and GC) x 4 (Test Moment: pretest, posttest, delayed posttest 1 and 2) mixed design, with repeated measures on the second factor. All participants completed a confirmation bias-pretest with hypothesis testing tasks, four-card selection tasks and illusory correlation tasks. Then they completed either an instruction in a domain-specific educational learning context with a video-based "consider-the-opposite" instruction with examples from the educational context, or an instruction in a domain-general learning context with relatively non-authentic general examples. Afterwards, all participants completed a 10-item topic interest questionnaire to measure an aspect of motivation, namely topic interest. Participants completed an immediate posttest, a delayed posttest after one week, and after five weeks, with isomorphic confirmation bias tasks.

In **Chapter 3**, we examined whether worked-example feedback on practice tasks during a "consider-the-opposite" instruction on reducing confirmation bias would lead to better performance on a confirmation bias-posttest and transfer test compared to correct-answer feedback on practice tasks, or no feedback. To the best of our knowledge, this was the first study to examine the effect of feedback on confirmation bias tasks and the first study to examine the differential effect of two types of feedback. Five conditions were tested: (1) a video-based "consider-the-opposite" instruction, practice tasks with elaborative feedback in the form of worked examples (WE, worked example), (2) a video-based "consider-the-opposite" instruction, practice tasks with correct answer feedback (CA, correct-answer), (3) a video-based "consider-the-opposite" instruction and practising without feedback (PO, practice only), (4) a video-based "consider-the-opposite" instruction, without practice (VO, video only) and (5) testing only (NT, no treatment). Conditions 3, 4 and 5 were created as control conditions to determine the learning effects of instruction and practice without feedback, instruction only and effects of pretesting without an intervention, respectively. We hypothesized that the order of performance gains and transfer of learning was: WE > CA > PO > VO > NT. The study was conducted amongst 132 psychology students and student teachers. For the learning measure, the experiment used a 5 (Condition: WE, CA, PO, VO, and NT) x 2 (Test Moment: pre vs post) mixed design with repeated measures on the second factor. For the transfer measure, the experiment was a single-factor (WE, CA, PO, VO, and NT) between-subjects design. To determine whether a learning effect (i.e., reduced confirmation bias from pretest to posttest) was due to the feedback, we also included the three control conditions. All participants first completed a confirmation bias-pretest with hypothesis testing tasks and four-card selection tasks. Except for the no treatment control condition, the other four conditions then watched the "consider-the-opposite" instructional video. Except for the no treatment control condition and the condition who only watched the instructional video, the three conditions practised with confirmation bias tasks. In both feedback conditions, participants either received elaborative, worked example feedback, or correct answer feedback on the practice tasks. In both feedback conditions and in the practice only condition, participants then self-reported their mental effort. Finally, all participants completed a confirmation bias-posttest and the transfer test.

The aim of the experiments reported in Chapter 4 and 5, was to examine whether learning by teaching on video was beneficial for student teachers' learning to reduce their confirmation bias (**Chapter 4**) and, in addition, to apply "consider-the-opposite" to design a citizenship education lesson plan in which perspective taking and open-mindedness play an important role (**Chapter 5**).

In **Chapter 4**, an experiment is presented in which we aimed to examine the effect of different generative learning strategies in combination with a "consider-the-opposite" instruction on student teachers' confirmation bias and a transfer task. Three conditions were tested: One group processed the instruction through preparing a lesson with the goal to subsequently teach the lesson on video (TOV, teaching on video), and one group through preparing a lesson only (PTT, preparing to teach). In the control group, participants were left to their own devices how to process the instruction that was provided to them on paper (CC, control condition). For learning and transfer it was hypothesized that TOV > PTT > CC. The experiment used a 2 (Confirmation bias-pretest vs posttest) x 3 (Instructional Strategy: TOV vs. PTT vs. CC) mixed design with repeated measures on the first factor. For the transfer test, we used a single-factor between subjects approach (i.e., instructional strategy: TOV vs. PTT vs. CC). As measurements for learning, the lesson preparations were analysed and together with the confirmation bias-posttest and transfer task performance it provided a broad view on the effectiveness of the different learning strategies. Additionally, questionnaires to measure feelings of social presence and arousal were designed and tested to explain any differences between the generative learning strategies.

The experiment presented in **Chapter 5** aimed to conceptually replicate the study presented in Chapter 4. Instead of the "consider-the-opposite" instruction, in this experiment an instruction that focused on how to prepare an open-minded lesson within citizenship education based on the principles of "consider-the-opposite" was provided to the participants in TOV, PTT, or CC. For this, awareness and knowledge on confirmation bias and perspective taking are necessary. This experiment used a single-factor between-subjects approach (i.e., instructional strategy: TOV, PTT, and CC) for the quality of a lesson plan on a social topic which served as the posttest. The quality of the explanation of the instruction, social presence, arousal (for TOV and PTT only), open-mindedness as measured by the Actively Open-minded Thinking scale, and the conceptual knowledge of the instruction were also measured. As in the previous study, it was expected that TOV > PTT > CC for the quality of the lesson plan and conceptual knowledge.

In **Chapter 6**, a summary of the main findings, methodological considerations, a discussion of the theoretical and practical implications, and directions for future research are presented.

Chapter 2

The effects of learning context on student teachers' confirmation bias and topic interest

This chapter has been submitted as:

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (submitted). "Consider the Opposite"– The effects of learning context on student teachers' confirmation bias and topic Interest.

Abstract

Confirmation bias refers to the tendency to selectively attend to and interpret information that is consistent with held beliefs or expectations, or to avoid information that is inconsistent with one's own preferences. In primary education it is important that teachers learn how to mitigate the confirmation bias because it might negatively affect decision making about their pupils and lessons. In the present study, a confirmation bias reducing "consider-the-opposite" instruction for student teachers was investigated under educational-practice or general-practice conditions. We hypothesized that practice in an educational context would reduce student teachers' confirmation bias more than practice in a general context because an authentic learning context has been shown to spark interest more. Results revealed that confirmation bias decreased equally in both learning contexts. However, participants reported more topic interest in applying "consider-the-opposite" in the educational context. Our results suggest that instruction and practice in an educational context motivates student teachers to learn how to reduce confirmation bias. It is suggested to include "consider-the-opposite" in the teacher education curriculum to support student teachers' unbiased decision making.

Keywords: confirmation bias, learning context, domain-specific instruction, consider-the-opposite, topic interest

Introduction

Perspective taking, i.e., considering evidence that is inconsistent with one's initial beliefs, is important to reduce the confirmation bias (Huang et al., 2016). This bias refers to the tendency to selectively attend to and interpret information that is consistent with held beliefs or expectations, or to avoid information that is inconsistent with one's own preferences (Baron, 1995; Nickerson, 1998). One way to enhance perspective taking and to mitigate the confirmation bias is by asking people to consider opposites and alternatives regarding their initial point of view (e.g., Larrick, 2004; Lord et al., 1984; Mussweiler et al., 2000). The thinking strategy "consider-the-opposite" (COS) consists of asking oneself the question, "What are some reasons that my initial judgment might be wrong?" (Larrick, 2004, p. 323). Hence, it encourages to direct attention to alternative evidence that would otherwise not be considered.

Studies have shown that a COS instruction reduces the confirmation bias (Lord et al., 1984; Schwind et al., 2012; Van Brussel et al., 2020). Lord et al. (1984) showed that participants reduced their confirmation bias by considering opposites immediately after an intervention in which they received explicit instructions to consider opposites after reading two research articles on the death penalty. Participants in the control condition were asked to be as unbiased as possible in evaluating the studies. One article explicitly supported capital punishment and the other article explicitly was opposed to it. Afterwards, participants were asked to indicate to what extent their attitude towards the death penalty had changed and to rate how convincing each study seemed as supportive or opposed evidence. Results showed that participants in the COS condition displayed less attitude polarisation on the immediate posttest and the authors concluded that biases in social judgment can be corrected by presenting evidence from both sides of an issue.

In the context of education, the confirmation bias might result in one-sidedness in the classroom. For example, a teachers' confirmation bias might lead to flawed reasoning about and acting differently towards pupils with the same capacities, which might differentially affect a pupil's academic progress (e.g., Darley & Gross, 1983; Rosenthal & Jacobson, 1968; Van den Bergh et al., 2010). Hence, it is important for teachers to learn how to reduce their confirmation bias through considering opposite evidence that they would otherwise not consider.

Several studies have shown that explicit instruction with practice, with or without feedback reduces biases (e.g., Adame, 2016; Heijltjes, 2014; Lee et al., 2016; Van

Brussel et al., 2020). An ongoing debate is whether critical thinking skills such as perspective taking should be taught as a separate course based on the idea that the skills are general and transferable to any discipline, or within the context of a specific discipline (Moore, 2004). In the domain-general approach, the instruction is taught without specific subject matter content: the to-be learned skill is the objective (Abrami et al., 2008; Ennis, 1989). According to Ennis (1989), the skills may be applied to different domains and domain-general instruction works therefore best. McPeck (1990), however, argued that thinking is always about something, and inherent, a domain-specific instruction works best. In short, for the present study we define a domain-general context as presenting general problems in a COS instruction and practice (i.e., the general context), and a domain-specific context as presenting the problems in an educational context (i.e., the educational context).

If instruction and practice is provided within the context of a specific subject, positive performance outcomes on e.g., reasoning and decision making are shown (Bangert-Drowns & Bankert, 1990; Heijltjes, Van Gog, & Paas, 2014; Marin & Halpern, 2011; Niu et al., 2013). For example, a general instruction in combination with economical practice tasks on decision making, enhanced students' reasoning skills on the short and long term (Heijltjes, Van Gog, & Paas, 2014). Furthermore, domain-specific problems enhanced physics students' reasoning and argumentation skills: the learning context made sense to learners, they showed interest in, and made them want to learn more (Tiruneh et al., 2016). However, to the best of our knowledge, learning context has not been manipulated in COS studies. Yet, there are reasons to expect that learning in a domain-specific context might have a positive effect on aspects of motivation and learning regarding COS.

A domain-specific context is aimed at constructing concepts and relationships that are situated in real-world problems that are perceived as relevant by the learner (Abrami et al., 2015; Brandsford et al., 1999). Abrami and colleagues (2015) showed in their meta-analysis that a domain-specific instruction has a positive effect on learning critical thinking skills; Presumably, because it helps a learner to understand the significance of an idea in the context of problems that are genuine to the learner. In addition, practising with domain-specific problems might enhance transfer of knowledge and skills to future job contexts (Brown et al., 1989).

Domain-specific learning contexts might also stimulate topic interest. Topic interest can be described as the level of interest that is sparked by a specific topic (Ainley et al., 2002). This motivational factor refers to the short-term involvement and the expression of individual interest in a specific subject (i.e., related to prior knowledge and intrinsic motivation, Mason et al., 2008; Schraw et al., 1995). Research by Wijnia et al. (2014) on problem based learning, has shown that it is important to take interesting problems into consideration because topic interest is a predictor of students' motivation and performance. In addition, studies have shown a positive relation between topic interest, reading comprehension and conceptual change, because topic interest promotes attention, effort and willingness to persist in the task, which leads to more efficiently processing of the information in the text and integrating new learning content into a student's network of conceptions (Duncan & Pintrich, 1992; Hidi, 2001; Mason et al., 2008; Schraw et al., 1995).

The present study. All in all, it seems reasonable to assume that the effect of a COS treatment on the reduction of confirmation bias depends on the degree of domain-specificity of the instruction and the practice tasks. We investigated this issue with student teachers who received a COS instruction and practice in an educational context or a general context. We were particularly interested in two research questions: (1) What is the effect of a COS instruction and practice tasks in an educational context on reducing student teachers' confirmation bias compared to instruction and practice tasks in a general context, and (2) What is the effect of a COS instruction and practice tasks in an educational context on student teachers' topic interest compared to the instruction and practice tasks in a general context?

To answer these questions, we measured learning outcomes on confirmation bias-pretest to posttests (immediately, after one and after five weeks), and topic interest.

Hypotheses. Firstly, we hypothesized that student teachers who receive instruction and practice in an educational context would show larger performance gains (i.e., reduction of confirmation bias) compared to a general context because learning in a relevant context has a beneficial effect on critical thinking performance (Abrami et al., 2015; Hill, 2000). Secondly, we hypothesized that participants in the educational context would report higher topic interest than participants in the general context because it will be perceived as more relevant than the general context (Duncan & Pintrich, 1992; Hidi, 2001; Mason et al., 2008; Schraw et al., 1995).

Method

Participants and design. Participants were 70 second year student teachers (M_{age} = 19.7, SD = 1.4; 59 women) from a Dutch teacher education institution. All participants gave informed consent prior to participating.

Reducing confirmation bias had not yet been taught in the curriculum. Participants were randomly assigned to the Educational Context (EC), or the General Context (GC) condition. The experiment used a 2 (Context: EC and GC) x 4 (Test Moment: Pretest, Posttest, Delayed Posttest 1 and 2) mixed design, with repeated measures on the second factor. The dependent variables were the scores on the confirmation bias tests with a higher score indicating reduced confirmation bias. After the practice tasks, topic interest was measured, and we compared the two between-subjects conditions on this dependent variable as well.

With G*Power (Faul et al., 2007), we calculated how many participants were required for a standard sensitivity of the test procedure for the Context x Test moment interaction effect on performance (i.e., power) of .80, a correlation between measures of .05 based on earlier conducted and similar experiments by the authors. With the final 44 participants who completed all four tests, we had sufficient power to detect a medium effect size of $f = .25$.

Materials. All materials described below (available upon request), were delivered online in Dutch through the Qualtrics platform (Qualtrics, 2017).

Instruction. An instructional video was shown in both contexts immediately after the pretest (see Figure 1). In the four-minute instructional video, a teacher explained what the confirmation bias is and the importance of perspective taking to reduce this bias. The teacher then explained by an example of a hypothesis testing task, how confirmation bias emerges in information searching. Next, a step-by-step explanation was given of the COS thinking strategy to reduce confirmation bias during the information search process. Both videos differed in the context of the examples in which the instruction was embedded. In the educational context, participants were addressed to as teachers, and in the general context they were addressed to as professionals at applied university level. In the educational context video, confirmation bias examples from the primary school context were used (e.g., “You are more likely to ask a child with highly educated parents questions to which you expect a more complex answer than a student from less educated parents and which you assume is less smart for that reason”), and in the general context confirmation bias examples in a general context were used (e.g., “You assume that a certain management training is very effective because many colleagues were enthusiastic. If you value their opinion, you are less inclined to look further for a more effective training for your development”).

Figure 1. QR-code of the instructional video (in Dutch).

Test and practice tasks. There were six tasks on the pretest, three practice tasks and seven tasks on the immediate posttest. Both delayed posttests consisted of six tasks. All tasks were hypothesis testing tasks: four-card selection tasks (Wason, 1968), illusory correlation tasks (Chapman, 1967), and hypotheses testing tasks based on a study by Jonas et al. (2001).

Wason's four-card selection tasks are often used in experiments to measure logical reasoning and confirmation bias (Evans et al., 1993; Heijltjes, Van Gog, Leppink, et al., 2014; Sellier et al., 2019; Stanovich & West, 2008; Van Brussel et al., 2020; Van Peppen, 2020). In these tasks, people need to solve a logic puzzle with four cards. See Appendix A for a task example. Participants had to decide which card(s) they needed to turn to test the truth of a proposition. Wason found that people in general select the card(s) that confirm(s) the rule and seldomly select a card that might disconfirm the rule. Consider, for example, the following rule: "If a card has a vowel on one side, then it has an even number on the other side. Which card(s) do you need to turn over to determine if the rule is true or false?". The four cards that are shown are A, D, 4, and 7. The only possible way to falsify the rule is by turning an affirmative card (A in the example) and by turning a non-affirmative card (7 = "uneven number" in the example). These cards are worth checking to test the rule. However, participants usually choose the cards that confirm the rule rather than cards that disconfirm, which is an expression of confirmation bias.

In the illusory correlation tasks (Chapman, 1967), a hypothesis was presented with a cross table in which research results about the hypothesis were presented. See Appendix B for a task example. Participants were asked whether the hypothesis was confirmed based on the research results. In this task, people have the tendency to only look for the confirming cells in the cross table. This is regarded as confirmation bias. To judge the hypothesis in these tasks well, one must consider all cells of the cross table.

The hypothesis testing tasks based on Jonas et al. (2001), showed how the biased search for information is related to someone's position on a statement. See

Appendix C for a task example. In this task, participants were asked whether they agreed with a statement (e.g., “Nowadays, children have less perseverance than children from the past”). After this, participants read eight short article summaries about this subject. Four articles were strongly in favour of and four were strongly opposed to the statement. Based on the given summaries, participants had to choose four out of the eight provided titles of articles they would like to read to gather more information about the statement. After selecting by ticking four boxes, the task was finished. Note that participants did not actually receive the articles. Participants showed confirmation bias by choosing mainly affirming articles when they agreed on the statement, or by choosing mainly disconfirming articles when they disagreed on the statement.

Topic interest questionnaire. Topic interest was measured after the practice phase by a 10-item scale that was developed by the authors. The questionnaire was based on the topic interest scales by Mason et al. (2008), and Schraw et al. (1995). Participants were asked to rate on a 5-point scale ranging from 1 (not at all) to 5 (very true) to what degree the statements applied to them. In the present sample, the 10 items had a Cronbach's $\alpha = .79$. In Table 2 the items are presented.

Procedure. After the pretest, and prior to the experiment, participants were automatically randomly assigned to one of the conditions through Qualtrics. Participants were seated in a computer room with dividers between each desk. The first author provided a short general verbal instruction about the nature of the experiment and some rules (e.g., it is not permitted to leave the lab during the experiment). They were instructed that this experiment was conducted to improve the curriculum. Headphones, pens and papers were provided, and participants started by clicking on a link that led to Qualtrics. Each phase (i.e., test, instruction or practice), was introduced separately before it started (i.e., the nature and number of tasks that had to be completed). Time was automatically logged for exploratory reasons. See Figure 2 for an overview of the procedure.

First, on the introductory page the procedure was explained, and participants filled in their name, age and prior education. Secondly, participants completed the pretest. Thirdly, participants were shown the instructional video and after that they practised three tasks in the context they were assigned to (EC or GC). Subsequently, they completed the questionnaire. Finally, all participants completed the immediate posttest. The delayed posttests took place one week and five weeks later.

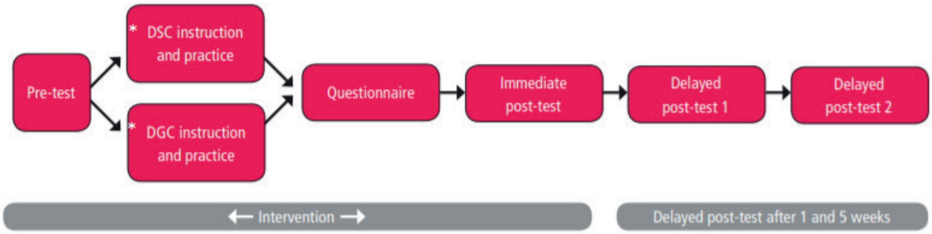

Figure 2. Overview of the experimental procedure (* DSC = domain-specific educational context and DGC = domain-general context)

Data analysis. For a correct answer on the four-card selection tasks and the illusory correlation tasks, participants received one point. For the hypothesis testing task, participants who chose five pieces of information that were in favour of their statement, received no points. If they chose one, two or three pieces of information in favour and respectively four, three or two pieces of information that disagreed with their statement, participants received two points because that showed a tendency for falsification. If they chose four in favour and one not in favour, they earned one point because that still showed too much confirmation bias.

For the pretest and both delayed posttests, participants could score a maximum of seven points, and eight points for the immediate posttest. The scores on the immediate posttest were recalculated (dividing them by 8, times 7), to be able to compare all tests.

Results

The pretest and immediate posttest ($n = 70$) differed in the number of participants compared to the first delayed posttest ($n = 62$) and the second delayed posttest ($n = 44$). No participants were excluded from the analyses. In all analyses reported below, a significance level of .05 was used. Due to the low Cronbach's α of the pretest ($\alpha = 0.11$) and the immediate posttest ($\alpha = 0.40$), the analysis of these test results was discarded[1]. Therefore, instead of using individual pretest results, we used the

1 Based on the correlation matrix of the posttest items, a subset of items was found with a higher Cronbach's $\alpha = .50$. (2, 4 and 7). However, an explorative ANOVA of the sum score of the subset revealed also no main effect of Context, $F(1, 68) = .19$, $p > 0.05$.

mean score on the pretest as a test value to determine reduction of the confirmation bias on the delayed posttests. The delayed posttests after one and five weeks showed Cronbach's α of .52 and .51, respectively.

Results for learning. Table 1 presents the confirmation bias test scores and SDs as a function of context and test moment.

Table 1. Pretest test value, Average Mean (M) Test Scores on the First and Second Delayed Posttests, Standard Deviation (SD), and Sample Size (n) as a Function of Context and Test Moment

	EC[a]		GC[a]		Total Mean	
Test moment	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Pretest test value	37	2.70 (1.04)	33	2.70 (1.04)	70	2.70 (1.04)
Delayed posttest 1	30	3.39 (1.19)	31	3.67 (1.46)	61	3.52 (1.32)
Delayed posttest 2	23	2.35 (1.03)	21	2.71 (1.23)	44	2.52 (1.13)

[a] EC = educational learning context, GC = general learning context

A One-Sample t-test showed that performance scores on the first delayed posttest were significantly higher than the mean average score on the pretest (M used as test value), $t(60) = 4.38$, $p < .05$. Subsequently, and in contrast to the planned analyses, a 2 (Context: EC vs. GC) x 2 (Test Moment: Delayed Posttest 1 and Delayed Posttest 2) mixed ANOVA with repeated measures on the second factor was conducted. There was no significant main effect of Context: $F(1, 42) = 1.10$, $p =.30$, with a very small effect size $\eta_p^2 = .03$. There was a significant main effect of Test Moment: $F(1,42) = 22.42$, $p < 0.5$, with a large effect size $\eta_p^2 = .35$; collapsed across conditions mean scores were higher on the first delay test than on the second. Furthermore, we did not find an interaction effect between Context and Test moment: $F(1,42) = 0.50$, $p = .483$, $\eta_p^2 < .01$. All in all, these results indicate that for both groups confirmation bias was reduced to a similar extent from the pretest to the first delayed posttest. In addition, confirmation bias increased again from the first delayed posttest to the second, with average confirmation bias scores returning to the pretest test value.

Results for topic interest. The ten items of the questionnaire were subjected to principal components analysis with Varimax rotation and Kaiser normalization to reduce the amount of variables and to determine coherent variable components. Items 5, 6 and 8 were recoded. After an initial analysis, items 6 and 8 were discarded, because the factor loadings were low. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .79 ("meritorious" according to Kaiser and Rice (1974)). Except for item 4, all KMO-values for individual items were greater than .5 which is the acceptable limit (Kaiser & Rice, 1974). Still, we included this item in the analysis because with .48 the KMO-value was so close to the acceptable limit that KMO-values without this item would not differ. The analysis was run to obtain eigenvalues for each factor in the data. Two factors had eigenvalues over Kaiser's criterion of 1 and in combination explained 59.44% of the variance. Table 2 shows the factor loadings after rotation. The items 7, 4, 3, 10 and 9 cluster on the same factor representing "topic interest of COS for classroom practice". This subscale had an acceptable reliability, Cronbach's $\alpha = .76$. The second factor consisted of the items 2, 5 and 1 and represents "topic interest of the lesson: instructional video plus practice". This subscale also had an acceptable reliability with Cronbach's $\alpha = .70$.

Topic interest of applying COS in the classroom. To test the effect of the learning context on the topic interest of applying "consider-the-opposite" in the classroom, an independent samples t-test was conducted. Therefore, we used each participant's mean score of the items 7, 4, 3, 10 and 9 from the questionnaire. The EC participants reported higher topic interest ($M = 5.55$, $SD = .72$) compared to the GC participants ($M = 5.12$, $SD = .70$): $t(68) = 2.56$, $p < .05$ with a medium of effect size of $d = .60$.

Topic interest of the lesson: instructional video and practice. An independent samples t-test was performed to compare the two learning contexts for the second factor. Therefore, we used each participant's mean score of the items 2, 5 and 1 from the questionnaire. There was a significant effect of context in favour of the EC participants ($M = 5.63$, $SD = .83$) compared to the GC participants ($M = 5.16$, $SD = .92$); $t(68) = 2.25$, $p < .05$ with a medium effect size of $d = .53$.

The results of both t-tests suggest that learning within the educational context might be more motivating for student teachers than learning in the general context.

Exploratory results on time spent on the intervention. We wanted to be able to control for time-on-task effects and explore whether the time spent on the intervention, might explain possible differences between the conditions on learning. Participants took an average 33.72 min ($SD = 8.61$) to complete all phases of the experiment. Both delayed posttests were left out of this analysis. It took EC participants significantly longer to complete the experiment ($M = 37.88$, $SD = 7.00$) compared to the GC participants ($M = 29.05$, $SD = 7.88$), $t(68) = 4.97$, $p < .05$ with a large effect size of $d = 1.18$.

Table 2. Factor Loadings for Factor Analysis after Varimax Rotation and Kaiser Normalization for the Topic Interest Questionnaire (n = 70)

	Rotated Factor Loadings	
Items	**Topic interest of instructional video**	**Topic interest of the lesson: COS and practice for classroom practice**
3: I think everyone needs to know more about confirmation bias and its influence on decision making	.72	.31
4: I want to learn more about "consider-the-opposite"	.76	.26
7: I think that during the lessons at the teacher training institute attention should be paid to confirmation bias	.73	.06
9: During my internship I will apply "consider-the-opposite" in my lessons	.62	.36
10: I would like to watch the instructional video again	.67	-.04
1: The assignments about critical thinking were very interesting for my future work as a teacher	.42	.68
2: It was easy to keep my attention on the instruction and the assignments about critical thinking	.07	.83
5: I did not find the assignments fascinating (recoded)	.17	.71

Note. Factor loadings > .60 are in boldface.

Discussion and Conclusion

The aim of the current study was to examine the effect of learning context on reducing student teachers' confirmation bias and to examine differences between two contexts on topic interest. We expected that a COS instruction within an educational context would lead to a stronger reduction of confirmation bias and that participants would report higher topic interest compared to the general instruction and practice tasks.

Both groups showed less confirmation bias on the first delayed posttest compared to the pretest. This result is consistent with earlier COS studies in which instruction in combination with practice reduced confirmation bias (Adame, 2016; Hirt & Markman, 1995; Lord et al., 1984; Mussweiler et al., 2000; Van Brussel et al., 2020). However, the present study also showed that the reduction was rather short-lived in both context conditions. Perhaps, participants might have used superficial strategies (e.g., heuristics) to solve the tasks immediately after the learning phase, which might explain performance gains between the pretest and the first delayed test (e.g., Dunlosky et al., 2013; Fiorella & Mayer, 2016). Perhaps, participants might not have been able to remember their rule-of-thumb for a long time, making it hard for them to sustain their performance over five weeks. For future research, it might be interesting to investigate why the effect of the instruction and practice was rather short-lived. For example, from our research we do not know which thinking strategy the student teachers actually used. Therefore, a qualitative follow-up study might be conducted to investigate how they actually solve the tasks from the present study. Should they indeed use rather superficial problem-solving techniques, then it might be worthwhile to examine which condition(s) should be created to stimulate deeper level understanding.

Furthermore, contrary to our expectations, we found that the reduction of confirmation bias from the pretest to the first delayed posttest was similar in both learning context conditions. At the same time, we did find robust effects of learning context on aspects of motivation as measured by the topic interest questionnaire. That is, student teachers in the educational context reported higher topic interest, and spent more time on task than participants in the general context condition. This lack of alignment between motivation and performance (i.e., reduction in confirmation bias) is in line with other studies from the literature, which show the complex relationship between motivation and learning (e.g., McConney et al., 2014; Wijnia et al., 2014). Lorencová et al. (2019) suggested that next to motivation and duration of practice, feedback has an important influence on critical thinking. It might be possible that indirect effects of motivation did not materialise in the present study because the intervention was relatively short, or no feedback was provided. This could be explained as follows: When instruction and practice is short, absolute differences between conditions in motivation related effort investment and/or depth of processing are restricted and that might have resulted in hardly noticeable performance differences. However, if the duration of instruction and practice would increase, these effort investment and processing effects become larger and might eventually result in a performance difference between a high-motivation and a low-motivation condition. Furthermore, providing feedback might be especially advantageous for participants with a higher level of motivation as they might spent more time processing, which in turn would lead to better performance (e.g., Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2005). However, in the current study, participants did not receive feedback on a practice task, and this may have diminished learning differences between the two contexts despite the higher level of motivation for the EC participants.

Although we used three types of confirmation bias tasks that were validated by previous biased reasoning research, combining these tasks in one test did not lead to one unidimensional construct to measure confirmation bias. We might therefore assume that participants, based on their low prior knowledge, resorted to response guessing. In turn, this led to a high degree of random fluctuation of participants' test scores, in particular of the pretest and immediate posttest scores. For future research, it is advised to provide feedback on practice tasks to prevent participants from response guessing. Also, to enhance student teachers' understanding when to apply COS in situations where the confirmation bias is at risk, other types of tasks that are applied to the educational setting might be better (Abrami et al., 2015; Lai, 2011).

Significance of the study. The results of our study contribute to the growing body of empirical knowledge on teaching COS as a critical thinking strategy to reduce the confirmation bias. Our results show that with a relatively short intervention, student teachers have learned to reduce their confirmation bias.

Student teachers reported higher topic interest by means of applying COS for classroom practice than student teachers who learned in a more general context. These promising effects call for further research on the role of motivational aspects in considering the opposite, next to cognitive aspects.

The relevance of our study for everyday practice, is the potential of (student) teachers to improve their reasoning skills by mitigating their confirmation bias. Suboptimal reasoning can have detrimental effects on pupils' academic achievement because it might lead to biased observations and subsequent incorrect decisions by the teachers (e.g., Darley & Gross, 1983; Rosenthal & Jacobson, 1968; Van den Bergh et al., 2010). In addition, an increasing amount of information, including false, misleading, or inaccurate information that is intentionally distributed to confuse or manipulate people is coming towards us via various sources (UNESCO, 2018). Therefore, people must be willing and able to consider opposing information and perspectives and give their beliefs and values a serious impartial consideration by setting aside one's commitment towards one's own perspective (i.e., being open-minded, Baehr, 2011). Hence, if people change their thinking strategies, and for example learn to COS, they learn to refrain from judgmental errors. Especially when stakes are high, people must be conscious of the confirmation bias by keeping an open mind to avoid poor decision making. Therefore, it is suggested to include perspective taking in the teacher education curriculum, because then the teacher is able to teach pupils in the classroom about critical thinking and unbiased reasoning, and how to transfer to new contexts, which in turn leads to better academic achievement, and decision making professionals in the future (Abrami et al., 2015; Lai, 2011; Nickerson, 1998; Rosenthal & Jacobson, 1968).

APPENDIX A

Example of Wason Four Card Selection Tasks

(translated from Dutch)

Educational context:

Candy

A teacher wants to know whether eating a lot of sugar rich candy causes ADHD. The rule is:

If a child eats a lot of candy, it is diagnosed with ADHD.
Which child(ren) should he examine to test his hypothesis?

- o A child that eats a lot of candy
- o An ADHD diagnosed child
- o A child that does not eat a lot of candy
- o A child with no ADHD diagnosis

General context:

Lease cars

An IT company employee makes an inventory of the car fleet. People with different functions drive cars from different price ranges. The rule is:

If one is consultant, then he/she drives a VW Passat.
Which card(s) should the employee turn to decide whether or not the rule is violated?

- o Seat Ibiza
- o IT-manager
- o IT-consultant
- o VW Passat

APPENDIX B

Example of Chapman's Illusory Correlation Task

(translated from Dutch)

Chinese or German?
A Dutch teacher would like to respond to the current situation in the world in which the economic and geo-political importance of China is increasing. Therefore, he teaches his pupils Chinese in fifth grade with the help of a native speaker.

His colleagues are less enthusiastic because Chinese is a very different to learn for native Dutch speakers. They assume that this makes it more difficult for pupils to learn. They advise their colleague to start teaching German in class because that language resembles Dutch more. They base their advice on a multilingualism study. Below is a table from that study. You see the number of participants, which language was learned as the second language and the results for Dutch. Based on the data from the table, check whether the hypothesis of the concerned colleagues can be accepted.

	Second language: German	**Second language: Chinese**
Increase of Dutch language results	180	140
Decrease of Dutch language results	80	60

APPENDIX C

Example of a Hypothesis Testing Task

(translated from Dutch)

Part 1
"Organic farming is better for people and the environment than regular farming".

Do you agree or disagree with this statement?

- I agree

- I disagree

Part 2
Information is available from articles on organic and regular farming. Below you will find summaries about the information that is available about this subject. You can assume that all sources are equally reliable. Firstly, read the information carefully and secondly, based on the summaries, choose four articles that you would like to read on this topic by ticking the box.

Example of two sources of which participants could choose from:

- In a journal on organic farming, it is written that food miles should be kept to a minimum, so it is important to buy meat from cows from the local barn and butcher. The meat is slaughtered by the butcher himself, so you know what you eat. It also tastes much better. Conclusion: organic farming is better for the environment than regular farming.

- In a farming journal, it is written that most food miles per product are made domestically. This applies, for example, to Dutch apples. They are harvested in late summer and stored in cold rooms for the rest of the year. After ten months in a cold store, they have used more energy than it takes to fly in fresh apples from New Zealand. Conclusion: regular farming is better than organic farming.

Chapter 3

Effects of elaborative feedback and correct answer feedback on reducing confirmation bias

This chapter has been published as:

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2020). "Consider the Opposite" Effects of elaborative feedback and correct answer feedback on reducing confirmation bias – A pre-registered study. *Contemporary Educational Psychology*, *60*, 101844. https://doi.org/https://doi.org/10.1016/j.cedpsych.2020.101844

Abstract

Unbiased reasoning is considered an essential critical thinking skill that students need to possess to face the future challenges in their work and life. Confirmation bias, which is the tendency to selectively attend to information that is consistent with held beliefs, presents a significant thread to unbiased reasoning. An effective strategy to reduce confirmation bias is the "consider-the-opposite"-strategy (COS). The central question of this pre-registered study was whether providing elaborative, worked example feedback after COS practice would lead to a better performance on previously practised and transfer tasks than correct-answer feedback. Participants were 132 university students who took a confirmation bias-pretest, watched an instructional video on COS afterwards and next received either worked example feedback or correct answer feedback on practice tasks, practised only, watched the instruction only or received no treatment. Finally, all participants took a learning test assessing their skill to avoid confirmation bias, and a transfer test assessing whether they could apply this acquired skill to problems containing other biases. Results revealed no differences on the learning test between both feedback conditions, but students who received feedback scored significantly higher on the confirmation bias problems than students who did not receive feedback. We carried out our pre-registered analysis plan, but due to the low reliability of particularly the pretest, we carried out an additional exploratory analysis on subsets of posttest items and a subset of transfer test items. Results on learning revealed the same pattern as the planned analyses. However, we found no differences between any of the conditions on transfer.

Keywords: elaborative feedback; correct answer feedback; confirmation bias; instructional video; learning; transfer

Introduction

Fostering students' critical thinking (CT), supporting the application of their CT skills, and achieving transfer of these skills to situations outside of their learning context has become more and more a central objective of education because it is considered to be crucial for decision making, leadership, judgment, and professional success (e.g., Barnett & Ceci, 2002; Beaulac & Kenyon, 2014; Catapano et al., 2019; Halpern, 1998; Hattie et al., 1996; Johnson & Hamby, 2015; Lobato, 2006; Quatrociocchi et al., 2016; Yang & Chou, 2008). However, there is no "one size fits all" regarding the definition of CT. Many definitions have been proposed but the broadest definition has probably been compiled by the Delphi Panel (Facione, 1990). The panel, which consisted of 46 philosophical experts, characterized CT as purposeful, self-regulatory judgment, which results in six CT skills, namely interpretation, analysis, evaluation, inference, explanation and self-regulation, and 19 dispositions (Facione, 1990). In this study we focus on the evaluation skill, specifically directed towards the critical assessment of information during decision making and judgment. To critically assess information, one needs to apply the disposition of open mindedness regarding alternatives and opinions (Abrami et al., 2015, p. 306 for an overview of CT skills and dispositions).

According to Evans' dual processing theory (2003), to critically assess information people should use reflective Type 2 thinking processes instead of automatic Type 1 thinking processes. The dual processing theory indicates that thinking errors occur because of rapid Type 1 reasoning instead of the desired, reflective Type 2 reasoning. The latter reasoning processes are slower and draw more heavily on working memory capacity than Type 1 reasoning (Kahneman, 2011). Excessive reliance on Type 1 thinking processes prevents reflection and may lead to cognitive biases. Cognitive biases are the result of applying mental shortcuts (i.e., heuristics) during judgment and decision making, and they appear when individuals draw inferences or adopt beliefs where the evidence for doing so in a rationally sound manner is either insufficient or absent (Haselton et al., 2015). An example of a cognitive bias is the confirmation bias. The confirmation bias is the tendency of people to selectively attend to information that is consistent with held beliefs and to fail to appropriately consider alternatives to his/her held beliefs or judgments (e.g., Jonas et al., 2001; Nickerson, 1998; Quatrociocchi et al., 2016; Reeves, 2014; Schwind et al., 2012).

The confirmation bias can be observed amongst others in rule testing situations. A rule follows a logical structure: if we assume that A is true then B should follow. According to general principles of rule testing two conditions should be met. First, if A is instantiated then the consequence B should follow. This is a confirmatory test. The second requirement is checking whether A does not occur when the consequence B is not present. This is a dis-confirmatory test. Consistent with the confirmation bias, it has been shown that in rule testing situations, people perform the confirmatory test but not the dis-confirmatory one. Cognitive psychologist Peter Wason (1960) was one of the first to coin the term confirmation bias in the context of rule testing. In Wason's four-card selection task, people are presented with four cards with a number on one side and a colour on the other. For two cards, the numbers are displayed (e.g., 3 and 6) and for two the colour (red and blue). Subsequently, they have to test a given rule, for example if a card has an even number on one side then its opposite side is red, by choosing which of four cards they need to turn over. Typically, people perform a confirmatory test (i.e., turning the card with the even number), but not the dis-confirmatory test (i.e., turning the card that is not red). So, people showed the tendency to confirm rather than to falsify the rule, which is an expression of the confirmation bias. In the literature, Wason's four-card selection tasks are often used to measure confirmation bias.

Next to rule testing, the confirmation bias occurs also in situations in which people need to select information to come to a decision or make a judgment. Quattrociocchi et al. (2016), showed that e.g., Facebook users have a tendency to search for information that supports their preference on a subject, and to reject disproving information that undermines their preference. In his overview article, Nickerson (1998, pp. 192-194) appoints examples of the confirmation bias in real-life situations, e.g., when diagnosing in medicine, reasoning in science and during judicial reasoning. An example of the latter is when a judge needs to come to a conclusion, he/she needs an open mind regarding the presented evidence. Therefore, judges need to disconnect the process of gathering information from that of drawing conclusions from the evidence, and they need to refrain from forming a personal belief about what happened during the crime and interpret the evidence as such. The confirmation bias appears when new evidence is interpreted based on that personal belief, because judges, as well as people in general, are more likely to remember information that is consistent with their own beliefs than inconsistent information (Nickerson, 1998).

To refrain from biased reasoning, decision making, and judgment caused by the confirmation bias, it is important that people learn the CT skills that support reducing this bias. It is generally assumed that learning to think critically does not develop and transfer automatically as a by-product of learning, instead it requires explicit instruction with practice (Abrami et al., 2015; Abrami et al., 2008; Bangert-

Drowns & Bankert, 1990; Heijltjes, 2014; Marin & Halpern, 2011; Markovits & Brunet, 2012; Mehta & Al-Mahrooqi, 2015; Niu et al., 2013). With respect to reducing confirmation bias, an instructional strategy that has been shown to be effective is encouraging learners to generate counter explanations. This stimulates learners to consider alternative explanations, which leads to a more balanced and objective evaluation of the evidence that is needed for decision making or judgment (Hirt & Markman, 1995). An example of such a strategy is "consider-the-opposite" (COS; see e.g., Adame, 2016; Hirt & Markman, 1995; Lord et al., 1984; Mussweiler et al., 2000). The central question of COS is: "What are some reasons that my initial judgment might be wrong?" (Larrick, 2004, p. 323). According to Larrick (2004), and Danielson and Sinatra (2017), COS works because this strategy encourages people to direct their attention to opposite evidence that would not otherwise be considered. It is known that presenting opposing information prompts learners to reconsider their prior position. For example, from the relational reasoning literature (i.e., the ability to differentiate meaningful patterns to enhance learning) it is known that reasoning about opposition (e.g., antithetical reasoning) facilitates understanding argumentation, persuasion, and conceptual change through the use of refutational texts and graphics (Danielson & Sinatra, 2017; Grossnickle et al., 2016).

In their study, Lord et al. (1984) induced COS through either explicit instruction to consider opposites or through stimulus materials that made opposite possibilities more salient. Their experiment was based on the assumption that biased assimilation of new evidence occurs when opposite possibilities are overlooked. Participants who either supported or opposed capital punishment were asked to read two summaries of bogus research articles which were pro or contra the death penalty including the applied methodology for each study. After reading, participants had to indicate to what extent their attitudes and beliefs towards capital punishment had changed and to rate how convincing each study seemed as evidence on the issue. In the control condition, participants were asked to be as objective and unbiased as possible in evaluating the studies. In the COS condition, participants were asked whether they would have made the same high or low evaluation had exactly the same study produced results on the opposing side of the issue. By this, the authors reminded the participants that a different experimental design might have brought different supporting cognitions to mind. Results showed that participants in the COS condition displayed less attitude polarization on the immediate posttest and the authors concluded that biases in social judgment can be corrected only by a change in strategy.

In the study by Lord et al. (1984), posttest results showed that participants reduced their confirmation bias by considering opposites immediately after the intervention. However, performance on the longer term is also an important educational goal. COS learning effects on the longer term were shown by Morewedge et al. (2015). In their study, participants either played an interactive, educational game or watched an instructional video with the aim to reduce confirmation bias. The instructional video consisted of defining several biases, examples and suggestions for mitigating strategies such as considering alternatives. In the educational game, each player was asked to find a missing neighbour who has to clear his name of any criminal activity with help of the player. During the game, an expert explained each bias and gave examples, and players made judgments designed to test the degree to which they demonstrate confirmation bias. Participants assessed their degree of bias during each game level, and they performed practice judgments of confirmation bias on which they received immediate feedback. Both training methods reduced confirmation bias immediately and moreover, the effect maintained after eight weeks. However, the game debiased participants more effectively than the instructional video during a posttest and a follow-up test (Morewedge et al., 2015). Morewedge et al. (2015) provided immediate feedback to participants on their performance on practice tasks and their level of self-assessed confirmation bias in the game condition but not in the instructional video condition. In their study, it was not clear whether the educational game, the feedback or the combination of both led to the learning effects in Morewedge's study (2015), because the feedback was a confounding part of the instruction.

However, it is reasonable to assume that frequent and continuous feedback produces greater understanding, performance and application of the learned knowledge (Hattie, 2012; Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2005). Feedback is conceptualized as information provided by an agent (e.g., teacher, peer, book, parent, self, experiences) to modify one's willingness to invest effort, thinking or behaviour in order to improve learning (Hattie & Timperley, 2007; Shute, 2008). Based on a synthesis of 1200 meta-analyses relating to influences on student achievement, Hattie (2015) showed that almost every instructional intervention has a learning effect but feedback has a large effect on student achievement (combined $d = .73$). The type of feedback, however, generates different effects on learning (e.g., Butler et al., 2013; Roelle et al., 2017; Shute, 2008).

Regarding feedback type, one distinction that can be made is between correct answer feedback and elaborative feedback. Correct answer feedback implies that learners only receive the correct answer after they provided their own response. Alternatively, a learner might receive more elaborative feedback for example consisting of the correct answer as well as the solution steps that lead up to the answer. The first type of feedback can be considered as feedback at the task level, whereas the latter type can be considered as feedback at the task level and at the process level. A review article by Bangert-Drowns et al. (1991) has shown that the effect of correct answer feedback on learning is small at best, whereas elaborative feedback, in the form of explaining the answer, produces large effects on learning (but see a study of Van der Kleij et al. (2012), for contradicting results).

These findings on learning may be explained by a particular effect from the Cognitive Load Theory (CLT) literature (Paas, Renkl, et al., 2003; Sweller, 1988; Sweller et al., 1998, 2019), namely the worked example effect. This effect entails that studying worked examples during instruction produces better learning with less investment of time and mental effort than solving conventional problems because worked examples support learners to construct cognitive schemas of how to solve problems, which can guide their learning during subsequent practice and problem solving (Cooper & Sweller, 1987; Paas, 1992; Sweller & Cooper, 1985; Sweller et al., 1998, 2019; Van Gog & Rummel, 2010; Ward & Sweller, 1990). Worked examples can be regarded as a form of elaborative feedback because a step-by-step explanation and demonstration of a strategy to arrive at the correct solutions is presented john. In other words, worked examples focus on the processes that lead to the correct answer. In contrast, conventional problems are problems for which learners need to find the solutions themselves, and which are generally less effective regarding investment of time and mental effort (Cooper & Sweller, 1987).

Thus, based on the literature presented above, it is reasonable to assume that adding feedback to an instructional design will mitigate the confirmation bias, particularly when the feedback contains worked examples and when the final test contains task, which are similar to those practiced during instruction (i.e., when the final test taps onto learning). To the best of our knowledge, there are currently no confirmation bias studies that tested this assumption. However, doing so is important to unearth optimal instructional design principles that help to reduce confirmation bias and to subsequently enhance judgment and decision making. The first question in the present study is therefore whether elaborative feedback with worked examples will reduce confirmation bias more than correct answer feedback on tasks that have been encountered during previous practice. Furthermore, an important question in the literature on critical thinking instruction is which design features enhance transfer of learned skills and knowledge to related, but not practiced, new tasks. A recent study by Butler et al. (2013) suggest that the type of feedback might be important with that respect. In their first experiment, Butler and colleagues tested the hypothesis that the effect of elaborative (worked example) feedback compared to correct answer feedback depends on how learning is assessed. Participants who read a text passage with critical concepts were required to answer questions about the text and either received correct answer feedback[2], elaborative feedback[3], or no feedback on the initial test about the text passages. The final test contained questions that were repeated verbatim from the initial test (assessing learning) and additional new inference questions which measured transfer. When retention was assessed with repeated questions on the final test, elaborative feedback produced equivalent performance relative to correct answer feedback. The key finding by Butler et al. (2013) however, was that subjects who received elaborative feedback on the initial test outperformed the other participants when understanding was assessed by asking participants to transfer their knowledge to a new context. In the literature on critical thinking instruction, the differential effect of elaborative (worked example) feedback and correct answer feedback on reducing the confirmation bias on transfer test has not been examined. Therefore, the second question addressed in the present study is whether a difference between the two feedback types on transfer can be observed in confirmation bias instruction (through applying COS).

The present study[4]. The aim of the present study was to determine what information a feedback message should contain in order to be effective for reducing confirmation bias and transfer of this effect to other bias tasks. Based on previous empirical work (Butler et al., 2013; Van der Kleij et al., 2012) and theoretical considerations (Hattie & Timperley, 2007; Shute, 2008) it seemed reasonable to expect that providing feedback on practice tasks will increase the impact of a debiasing intervention on the reduction of confirmation bias and that providing elaborative feedback in the form of worked examples will be more effective than merely corrective feedback. The experiment was conducted to address this hypothesis.

Five conditions were tested: 1) an instructional COS video, practice tasks with elaborative feedback in the form of worked examples (WE), 2) an instructional COS video, practice tasks with correct answer feedback (CA), 3) an instructional COS video

---

2 A statement of the correct answer

3 The correct answer and in addition text passages that explained the concept but did not provide the answer to the inference question (transfer)

4 The present study was pre-registered at the Open Science Framework: hypotheses, planned data-collection, planned methods, and planned analyses can be found on https://osf.io/thn6x/?view_only=4e96d3e189744b39a0de82ee73131e52.

and practising without feedback (PO), 4) an instructional COS video, without practice (VO) and 5) testing only (NT). Conditions 4 and 5 were created as control conditions to determine, respectively, learning effects of instruction-only and effects of pretesting without an intervention. It was explored whether the form of feedback has an effect on self-reported mental effort during practise in the WE, CA and PO condition. Learning and transfer outcomes of bias tasks on which COS can be applied were measured to answer the research questions. Therefore, test outcomes on isomorphic test and practice confirmation bias tasks (i.e., same structure, different cover story) and transfer tasks (i.e., tasks on other biases) were used.

Hypotheses. In our design, all participants took an isomorphic pretest and posttest containing items that were addressed during instruction and that were practised in the WE, CA, and PO condition. The pre-post performance gain score was assumed to reflect learning. On the posttest, we also measured transfer in all five conditions. Based on the literature presented above, we hypothesized that the performance gain in the WE condition would be larger than in the four other conditions (Hypothesis 1a). Furthermore, it was hypothesized that performance gain in the CA condition would be larger in the PO, VO and NT condition, since feedback is more effective for learning compared to no feedback at all (Hypothesis 1b). Moreover, it was hypothesized that performance in the PO condition would be larger than the VO and NT condition because instruction is a key factor in enhancing CT skills (Hypothesis 1c). Finally, it was hypothesized that performance in the VO condition would be larger than in the NT condition, partly because participants were exposed to the instructional video (Hypothesis 1d).

Regarding transfer, we hypothesized that performance in the elaborative feedback (worked examples - WE) condition would be larger than in the four other conditions (Hypothesis 2a) because elaborative feedback enables learners to better comprehend the concepts and thus facilitates the application of that knowledge to new contexts. Finally, it was hypothesized that performance scores in the correct answer feedback condition (CA) would be larger than performance in the conditions without feedback on posttest transfer performance because correct answer feedback is more effective on transfer than no feedback at all (Hypothesis 2b). For the remaining conditions (i.e., PO, VO and NT) we expected the same ordering on transfer performance as on learning (hypothesis 2c).

In addition, mental effort investment was used to exploratively assess possible differences between types of feedback (WE and CA), and no feedback (PO). Therefore, mental effort during practise was taken into account as an index of the cognitive load (Paas, 1992). Mental effort is defined as "the total amount of controlled cognitive processing in which a subject is engaged" (Paas & Van Merriënboer, 1993, p. 738).

Method

Participants and design. Participants were 132 first year psychology students from a Dutch university (n = 107) and second year student teachers from a Dutch university of applied sciences (n = 25), who were randomly assigned to one of only five conditions (M_{age} =20.73 , SD = 3.46, 114 women). All subjects gave informed consent prior to participating in the present experiment. With regard to their prior education, 90% reported the required secondary school level to enter higher education or university in the Netherlands, and 10% reported other prior education, such as a completed study at a university of applied sciences. At the time of the experiment CT had not yet been taught in the curriculum. For the learning measure, the experiment used a 5 (Condition: WE, CA, PO, VO and NT) x 2 (Test Moment: pre vs post) mixed design with repeated measures on the second factor. For the transfer measure, the experiment was a single-factor (WE, CA, PO, VO and NT) between-subjects design. Figure 1 presents a schematic overview of the five conditions.

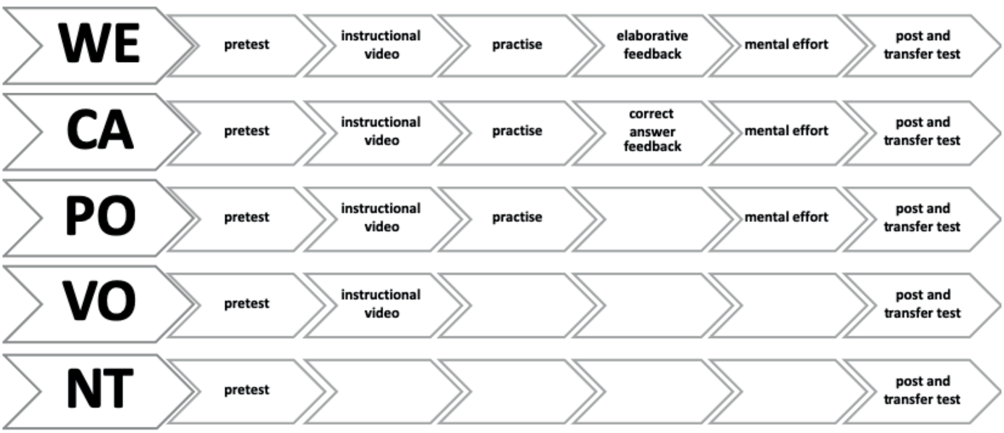

Figure 1. Schematic overview of the five conditions

We used G*Power (Faul et al., 2007) to determine the required sample size for a standard sensitivity of the test procedure (i.e., power) of .80 under a fixed α level of

.05, a correlation between repeated measures of 0.08 based on an earlier conducted and similar experiment by the same authors, and an estimated medium effect size of $f = .23$ (cf. Cohen's $d = .90$ by Butler et al., 2013). From the G*Power calculation a total sample size of 110 was required. For the transfer tasks, which had the same requirements as the confirmation bias-posttest tasks to achieve a power of .80, 235 participants were needed. Therefore, the aim was to test as many participants as possible within the available execution time with $n = 110$ as the lower limit.

Materials. All materials were delivered through the online Qualtrics platform and published on the Open Science Framework. The material was presented in Dutch.-

Instructional video. See Figure 2 for a QR-code that leads to the instruction on YouTube (in Dutch). In the 4.05-minute instructional video a teacher explained how confirmation bias is related to CT and the importance of reducing confirmation bias and the necessary CT skills. The teacher then explained how confirmation bias emerges in the information searching process related to a hypothesis testing task from the pretest. Next, a step-by-step explanation of COS (Lord et al., 1984) as a strategy to reduce confirmation bias during the information searching process was given. On the screen the steps were shown as text. The first step of COS that was mentioned was to ask oneself the question: "What are possible reasons that my initial judgment might be wrong?" as used in the original study by Lord et al. (1984). Next, one had to "consider the opposite" which meant that one must consider contradictive information. The last step in COS is that one has to critically assess the contradictive information and weigh both answers before making a final decision.

Figure 2. QR-code of the instructional video (in Dutch)

Test tasks. All tasks on the pretest and posttest were designed by the first author. The tasks were based on widely accepted tasks regarding measuring confirmation bias and included four-card selection tasks (Wason, 1968) and hypothesis testing tasks (Jonas et al., 2001). The pretest as well as the posttest consisted of four four-card selection tasks and two hypothesis testing tasks.

For the four-card selection tasks, participants had to choose which of the four given cards they needed to turn over in order to test a given rule. In general, participants have the tendency to select two affirmative cards, however, in order to test the rule, they need to turn over an affirmative card as well as a non-affirmative card. The tendency to only select the affirmative cards can be seen as an expression of confirmation bias (Nickerson, 1998). The Appendix shows and example of this task with both types of feedback. The rule is: "If the card shows an even number on one face, then its opposite is red". The four cards that are shown are "Red", "Yellow", "3", and "8". The only possible way to falsify the rule is by turning an affirmative card ("8" in the example) and by turning a non-affirmative card ("not red" = "yellow" in the example). These cards are worth checking to test the rule. However, participants usually choose the cards that confirm the rule rather than cards that disconfirm, which is an expression of confirmation bias.

The hypothesis testing tasks based on Jonas et al. (2001), showed how the biased search for information is related to someone's position on a statement. In this task, participants were asked whether or not they agreed with a statement (e.g., "Nowadays, children have less perseverance than children from the past") and were asked to motivate their answer to stimulate them to think seriously about their statement. After this, participants read eight short summaries of articles about this subject. Four articles were strongly in favour and four were strongly opposed to the statement. Based on the given summaries, participants had to choose four articles which they wanted to read to gather more information about the statement. After selecting, the task was finished. Note that participants did not actually receive the articles. Confirmation bias is expressed by participants if they choose information that agrees with their initial point of view. If their choice is more balanced (choosing e.g., two articles in favour and two articles contra), their confirmation bias is mitigated.

Practice tasks. These tasks were similar to the test tasks. Participants practised four tasks: three four-card selection tasks and one hypothesis testing task. After each practice task, participants reported experienced mental effort using the 9-point Mental Effort Rating Scale developed by Paas (1992).

Transfer test tasks. The six transfer tasks provoked other biases than confirmation bias although a person could apply COS as well to come to the correct answer. The tasks are discussed below and they are based on sample items from the Comprehensive Assessment of Rational Thinking (CART) by Stanovich et al. (2016). Also, the structure of the transfer tasks deviated from the instructed and practised tasks. Participants had to motivate their answer after each transfer task.

We used a task based on the research by Snyder and Swann (1978) on hypothesis testing which differed in structure from the practised tasks. In this task, participants were provided with information about extravert behaviour. They were told they had to choose 12 out of the 27 provided questions (11 on extraversion, 11 on introversion and five neutral questions) to interview a child to find out whether the child is an extravert. Confirmation bias is at risk here, because people have the tendency to choose questions that confirm the extraversion (e.g., “What would you do to enliven a boring party?”) and avoid asking questions that must prove the opposite (e.g., “Which things do you dislike about a noisy party?”). Two transfer tasks were based on conjunction fallacy tasks used by Tversky and Kahneman (1983) on the role of heuristics in decision making and the conflict between heuristics and logic. We used the “Linda”-conjunction fallacy task in which the probability of two events occurring together can never be larger than the probability of any of these events occurring alone. Three transfer tasks were included based on tasks by Stanovich et al. (2016). Two of these tasks presented problems that evoke a direct, intuitive but incorrect answer where reflection is more appropriate to solve the problem. Participants intuitively select an answer, which they assume is correct. However, in these tasks, the intuitive answer is typically incorrect and getting to the correct answer requires at least that people take a step back and think about why their initial answer might be incorrect. This is a form of considering the opposite. An example of such a task is: “The number of bacteria in a bin doubles every hour. If it takes 32 hours to completely fill the container, how much time do the bacteria need to fill half the bin?”. The intuitive answer is 16 but the right answer is 31. Finally, a base rate fallacy task was used in which probability is a problem (cf. the “Linda”-problem). This fallacy refers to the phenomenon whereby people ignore or undervalue probability, typically in less informative but more intuitively appealing information about a case (Kahneman & Tversky, 1973). For this, we translated a case by Stanovich et al. (2016, p. 333) in Dutch. In this case a professor must choose between two textbooks for a new course. Participants are being told how she thoughtfully comes to a choice, however at the end one less relevant and reliable piece of information for the other textbook is given. Participants are asked to advise the professor which one of the two textbooks she should definitely or probably must use.

Compared to the practice tasks, the items on the transfer test generally require a near transfer according to Barnett and Ceci’s taxonomy (2002). On the transfer tasks, participants had to apply COS to solve the problem. Furthermore, because the entire procedure took place in a single session without focus on a specific knowledge domain, the training tasks and the transfer test items show considerable overlap in physical and temporal context. The modality overlap was also substantial between the practice tasks and the transfer test items as both contained short, written cases.

Procedure. The experiment was conducted at a Dutch university and a Dutch university of applied sciences in a computer room with individual cubicles. By participating, university students earned course credits and students from the university of applied sciences participated voluntarily. The experimenter provided a short, general verbal instruction about the nature of the experiment and explained some rules (e.g., mobile phones must be switched off and it is not permitted to leave the lab during the experiment). Participants were asked to read the procedure and by clicking to continue, they agreed to participate in the experiment. Prior to the experiment, participants were randomly assigned to one of the five conditions. Headphones were provided and participants started by clicking on a link that lead to the Qualtrics platform (Qualtrics, 2017).

Participants were presented with each element of the allocated condition on a separate page on the computer screen. They could not switch to previous tasks, nor were they allowed to continue until the current task was completed. The experiment ended with a short demographic questionnaire (age, gender and prior education). After completing the test, participants were thanked and left the room. Total time was automatically logged by Qualtrics, and during the practice phase, all elements were logged separately.

As can be seen in Figure 1, participants in the WE condition received elaborative feedback after completing each practice task. The elaborative feedback consisted of a worked example. They had to read the text which consisted of a step-by-step elaboration of the task. In the CA condition, only the correct answer was provided as feedback after each practice tasks. The Appendix shows a four-card selection practice task with examples of both feedback types. The PO group practised after the instruction without receiving feedback. Participants in the VO conditions watched the instructional video after the pretest and started making the posttest immediately afterwards. In the NT condition, participants only took the pretest and the posttest.

Data analysis. All participants could score a maximum of eight points on the pretest. On the posttest participants could score a maximum of fourteen points: eight for the learning tasks and six for the transfer tasks. The data of all participants (n = 132) were included in the analyses. Results are reported with and without outliers

For each correct answer on a four-card selection task, participants received one point. The answer was correct when participants chose one card which confirmed the rule and one that disconfirmed the answer. For the hypothesis testing tasks based on Jonas et al. (2001), participants received one or two points, depending on the information which they selected: participants who selected four pieces of information that were either all pro or contra the statement, they received no points. If they selected two pieces of information in favour and two pieces of information that disagreed with their statement, participants received two points because that showed a tendency for falsification. If they selected three pieces of information in favour and one not in favour, they earned one point because that meant they still showed too much confirmation bias. The pretest with six items had a Cronbach's α = .16, so the proportion of systematic variance in the pretest sum score was low. This was most probably due to the low level of relevant prior task knowledge, resulting in response guessing and consequently a high degree of non-systematic variance in the pretest sum scores. Furthermore, the posttest with six items had a Cronbach's α = .58, which is still low for the purpose of the presents study, namely detecting differences between group means. We carried out our pre-registered analysis plan, but due to the low reliability of particularly the pretest, we carried out an additional exploratory analysis on the posttest items. Because the reliability of the posttest total score was suboptimal, we examined Cronbach's α for two subsets of posttest items, i.e., the two hypothesis testing tasks and the four-card selection tasks, for which the item responses can be expected to be correlated as the items in each subset are isomorphic. The two hypothesis testing tasks had a negative correlation with Cronbach's α = -.17. The other subset of the four-card selection tasks had an acceptable Cronbach's α = .76. Hence, for the posttest, we conducted an exploratory analysis on the sum score for the latter subset.

For the six transfer tasks, one point was assigned if one provided the correct answer and the correct explanation. If a participant gave the correct answer but an incorrect explanation or an incorrect explanation but a correct explanation then 0.5 points were assigned. If a participant gave an incorrect answer as well as an incorrect explanation then no points were assigned. The transfer test had a low reliability, Cronbach's α = .24. Thus, the total-test scores cannot be meaningfully interpreted. Hence, and therefore, contrary to what was mentioned in the pre-registrational part of the study, explorative analyses per item and on subset of items were conducted on the items from the transfer test.

Qualtrics automatically scored the responses and total time spent on the intervention. The scores were transferred to SPSS for the statistical analyses. The mental effort scores and the time measurement during the practice phase in the WE, CA and PO conditions, were computed.

Results

The data of all participants (n = 132) were included in the analyses. Results are reported with and without outliers (if necessary), and for planned as well as explorative analyses. In all analyses below, a significance level of .05 was used. Partial eta-squared (η_p^2) is reported as a measure of effect size for the ANOVAs for which 0.01 is considered small, 0.06 medium, and 0.14 large. See Table 1 for means and SD of pretest (eight points maximum), posttest without transfer scores (eight points maximum), self-reported mental effort means (nine points maximum), and time spent on practising. In the WE, CA, PO, and VO condition participants watched the 4.05 min instructional video. The means and SD of the explorative transfer results with a maximum of six points are presented in Table 2.

Planned analyses

Performance on learning. A 5 (Condition: WE, CA, PO, VO and NT) x 2 (Test Moment: pre vs post) mixed ANOVA on the learning scores with repeated measures on the second factor was performed. We found a main effect of Test Moment on learning, $F(1, 127) = 99.66$, $p < .001$, $\eta_p^2 = .44$. However, and most importantly, there was a Test Moment x Condition interaction effect on learning, $F(4, 127) = 10.29$, $p < .001$, $\eta_p^2 = .25$. As a follow up of this significant interaction, a planned Helmert contrast with NT as baseline was performed. As expected, this contrast showed that the mean learning gain in the NT condition was significantly lower with a contrast estimate of -1.09 (SE = .39), $p < .05$, than the combined mean learning from the subsequent four conditions. Furthermore, with a contrast estimate of -.81 (SE = 34), $p < .05$, the mean learning gain in the VO condition as expected was significantly lower than the combined mean learning gain in the PO, CA and WE condition. Also as expected, the mean learning gain in the PO condition was significantly lower than the combined mean in the two feedback conditions with a contrast estimate of -1.80 (SE = .37), $p < .05$. Contrary to what was expected, the two feedback conditions WE and CA did not differ in mean learning gain with a contrast estimate of -.29 (SE = .43), $p > .05$.

Table 1. Mean Scores and SD on Pretest, Posttest, Mental Effort Rating and Practice Time (min) per Condition

	WE *(n = 25)* *M (SD)*	CA *(n = 26)* *M (SD)*	PO *(n = 26)* *M (SD)*	VO *(n = 28)* *M (SD)*	NT *(n = 27)* *M (SD)*
Pretest	3.56 (0.25)	3.62 (1.27)	3.37 (0.97)	3.15 (1.54)	3.56 (0.64)
Posttest	6.04 (1.46)	5.81 (1.27)	3.89 (1.09)	4.11 (1.67)	4.00 (1.36)
Mental Effort	4.54 (1.83)	4.23 (1.56)	4.27 (1.37)		
Practice Time	4.39 (1.76)	5.06 (2.06)	4.80 (1.51)		

Five paired t-tests were conducted to evaluate the differences within the groups from pretest to posttest performance. There was a statistically significant increase from pretest to posttest for the WE group (M_{dif} = 2.48 , SD = 1.46), t(24) = -6.96, p < .001, CA group (M_{dif} = 2.19, SD = 1.27), t(25) = -7.466, p < .001, PO group (M_{dif} = .54, SD = 1.09), t(25) = -2.34, p = .028, and VO group (M_{dif} = .93, SD = 1.65), t(27) = -3.06, p = .005. There was no statistically significant increase for the NT condition (M_{dif} = .44, SD = 1.36), t(26) = -1.56, p = .130. Outliers were detected in the CA group (5 high extremes) and PO group (1 high extreme and 2 low extremes), but results were similar with outliers.

Mental effort scores. To gain more insight into the mean mental effort per group as a function of type of feedback, or no feedback, mental effort scores were collected after each of the four practised items and the feedback in WE and CA and after the four practised items in PO. Mean results per condition are shown in Table 1. The mean of the self-rated mental effort scores over all three conditions was 4.35 (SD = 1.59). The mental effort rating scale by Paas (1992) indicates this score as rather low mental effort. For this, a mixed ANOVA was performed. We found no significant effects of condition on mental effort, F(6, 148) = .736, p = .621, η_p^2 = .029.

Practice time and feedback. Practice time and time spent on reading feedback (WE and CA) and practice time (PO) was used for exploratory reasons only. We found no significant difference between the time participants spent on practising and reading the feedback in the WE, CA, and PO conditions: F(1, 49) = 1.53, p > .05, η_p^2 = .03.

Explorative analyses

Explorative analyses on learning (posttest performance). As mentioned above, beyond our pre-registration, we conducted analyses on a subset of similar posttest tasks, namely the four-card selection posttest tasks. For this subset, we found a significant effect on posttest performance between conditions, $F(4, 127) = 17{,}32$, $p < 0.05$, $\eta_p^2 = .35$. A Helmert contrast showed the same pattern as the pre-registered mixed ANOVA[5]. That is, the mean posttest performance in the NT condition was significantly lower with a contrast estimate of -.21 (SE = .07), $p < .05$, than the combined mean posttest performance from the subsequent four conditions. Furthermore, with a contrast estimate of -.24 (SE = .07), $p < .05$, the mean posttest performance in the VO condition was significantly lower than the combined mean posttest performance in the PO, CA and WE condition. Also, the mean posttest performance in the PO condition was significantly lower than the combined mean posttest performance condition in the two feedback conditions with a contrast estimate of -.50 (SE = .07), $p < .05$. The two feedback conditions WE and CA did not differ in mean posttest performance with a contrast estimate of -.02 (SE = .09), $p > .05$.

Explorative analyses on transfer. See Table 2 for the descriptive statistics: the mean scores and SDs of the subsets of transfer test items we used for the explorative analyses. Cronbach's α for the six transfer test items was very low ($\alpha = .24$), so we conducted some explorative analyses. Based on the correlation matrix of the transfer items, we tried to identify clusters of items with positive correlations. A subset of three items appeared to form a cluster (Cronbach's $\alpha = .54$). These items were the two reflection vs. intuition tasks by Stanovich et al. (2016) combined with one of the conjunction fallacy tasks (Tversky & Kahneman, 1983). An explorative ANOVA on the sum scores of these three items failed to reveal a significant effect of Condition and the effect size was small to medium ($\eta_p^2 = .05$). The three other items were analysed separately. Again, there were no significant results, and the effect sizes of these tasks were small to medium[6].

---

5 We also conducted an exploratory principal component analysis (PCA) on the six items of the posttest but the results did not deviate from the One-way ANOVA.

6 We also conducted an exploratory principal component analysis (PCA) on the six items of the transfer test but the results did not deviate from the exploratory One-way ANOVA.

Table 2. Mean Scores and SD of Subsets of Transfer Test Items

	WE	CA	PO	VO	NT
	(n = 25)	*(n = 26)*	*(n = 26)*	*(n = 28)*	*(n = 27)*
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
3 Item subset*	1.36 (1.07)	1.65 (1.06)	1.81 (0.94)	1.75 (1.08)	2.11 (1.09)
Base rate	0.56 (0.51)	0.52 (0.50)	0.50 (0.51)	0.50 (0.51)	0.74 (0.45)
Linda	0.40 (0.50)	0.19 (0.40)	0.27 (0.45)	0.46 (0.51)	0.30 (0.47)
Extravert	0.64 (0.49)	0.69 (0.47)	0.58 (0.50)	0.57 (0.50)	0.37 (0.49)

* Two reflection vs. intuition tasks by Stanovich et al. (2016) combined with a conjunction fallacy task (Tversky & Kahneman, 1983)

Conclusion and Discussion

The aim of the experiment was to examine the effect of elaborative feedback in the form of worked examples on reducing confirmation bias and transfer compared to correct answer feedback or no feedback. We hypothesized that providing elaborative feedback would be more effective than providing correct answer feedback, or no feedback. The results showed that both feedback types were beneficial for learning how to avoid confirmation bias in hypothesis testing tasks compared to practising without feedback, or no practising. This finding is in line with the findings by Butler et al. (2013), and corroborates the effects of feedback on learning that were reported in other domains.

The differential effect of feedback type on learning might be due to various reasons. The first reason might be related to the remarkable finding that we did not find any differences between the WE and the CA condition in time spent on practice and reading the feedback. This finding might have emerged because participants in the WE condition did not engage in a deep-level processing of the worked-out solutions. Participants might have focused on what the correct solution steps were but might not have focussed on why these solution steps were correct (e.g., Renkl & Atkinson, 2010). As a result, the benefit of WE feedback over CA feedback might have not been effectuated. Finding an explanation for the deviation of the results from the expectations might however be an example of hindsight bias (e.g., p-hacking and subsequent inflations of false positive rate (Simmons et al., 2011) and HARKing: hypothesizing after the results are known (Kerr, 1998)). To increase the credibility of our study and to prevent ourselves from our own cognitive biases, we pre-registered our research plan.

Secondly, the order in which the information was presented in the hypothesis testing tasks might also have influenced the WE-CA comparison. Concerning the order of presenting information, in their study, Jonas et al. (2001) presented the information to participants in either a sequential or simultaneous order. Sequentially presenting information caused a stronger preference for supporting information (i.e., confirmation bias) than simultaneous presenting. Our participants received the information on Jonas et al.'s hypothesis testing task on the posttest simultaneously, which might have resulted in lower levels of confirmation bias overall. In addition, it might be possible that a ceiling effect occurred because most participants probably found out that choosing balanced information is the principle, which made it hard to find differences between conditions.

According to Nickerson (1998), in general people do not perform well on Wason four-card selection tasks because its abstractness does not relate to everyday hypothesis testing. Many studies come to the same conclusion (e.g., Kellen & Klauer, 2019; Nickerson, 1998; Ragni et al., 2018; Van Peppen et al., 2018). In addition, the complexity of the four-card selection task lies in the fact that people over and over select the card(s) that confirm(s) the rule and seldomly select a card that might disconfirm the rule (Ragni et al., 2018; Wason, 1968). In other words, people often do not consider the opposite spontaneously but use heuristics that allow for fast information processing and that lead to thinking errors and biased reasoning when instead more reflective thinking must be applied (see Evans, 2003 on Type 1 and 2 processing). During the practice phase, participants in the WE and CA condition might have learned what the correct answers are to this kind of tasks given them an advantage on the posttest over the other conditions. However, due to the task complexity, participants in the WE condition might not have been able to fully understand the underlying rules of logic of the Wason four-card selection tasks they received in their feedback. As a result, this extra information could not be used to give them an advantage over the participants in the CA condition on the post test.

Participants did not receive customized feedback but general feedback, which differed in the degree of elaboration (the correct answer only or elaborative feedback). This might have caused superficial studying when a participant gives a correct answer, because it is then conceivable that a participant does not read the solution thoroughly. For future research, providing customized feedback on the given answer might be more effective: e.g., when a participant gives the correct answer, one could only give the elaboration on the solution steps without an explicit referral to the correct answer; this might prevent a superficial processing of the correct answer only. By contrast, in case of an incorrect answer more background information in combination with elaboration on the solution might enhance deeper learning. Also, exploring the learning effects of other types of feedback on confirmation bias tasks than the ones used in the present study might be considered in future research. Hattie and Timperley (2007, p. 88) for instance mention feed up ("Where am I going?") and feed forward ("Where to next?") next to feedback ("How am I doing?").

For the transfer outcome measure, we did not find any effects of the conditions in our experiment. This was most probably due to the low reliability of the transfer test. Hence, other transfer test items must be created if researchers plan a follow-up study along the lines of the present study. The number of test items based on the three transfer test items that showed a somewhat acceptable (but still low) reliability, can be extended for a transfer test. Interesting to mention is that the explorative ANOVA of the transfer test tasks based on Snyder and Swann (1978) revealed the same pattern as the learning phase although these findings were not significant. Although the difference was not significant, this pattern could indicate that participants who received feedback (WE and CA) were able to apply the learned knowledge to tasks that differ only slightly on structure compared to the learning tasks, because the Snyder and Swann-based task most closely resembled the hypothesis testing tasks from the instruction and learning phase. However, and also not significant, the reversed learning pattern was found on the base rate fallacy problem based on Stanovich et al. (2016), because participants who received no treatment (NT) scored the highest and participants who received worked example feedback (WE) scored the lowest on this task. Instruction, practise and feedback did probably not lead to the relevant knowledge structure to achieve transfer. Therefore, applying COS might have hindered participants in solving this task. However, as reliability was very low and results were not significant, the interpretation of these patterns must be exercised with great caution, but it is interesting for further research because transfer effects are difficult to establish in unbiased reasoning studies (e.g., Heijltjes, 2014; Van Peppen et al., 2018).

Apart from transfer, we did not find any effects of our conditions on mental effort and time spent on practising. These results might be explained by Evans' dual processing theory (2003). If our participants applied a heuristic, Type 1 thinking is sufficient for learning. Since our results on transfer were not significant, one might assume that participants struggled to apply the learned strategy to new contexts.

Limitations. The study has some limitations. First, participants were not asked to reason about their answers on the tasks. Research shows positive learning and transfer effects of e.g. self-explanation on tasks through which participants are asked to reason on their answer (Chi et al., 1994; Fonseca & Chi, 2013). However, other research did not show effect of self-explanation prompts on learning and transfer (Van Peppen et al., 2018). For further research, it is interesting to gain more insight in the reasoning process in order to enhance learning and transfer performance, taking e.g., insights from the relational reasoning literature into account that focus on the support of analogies, anomalies, antinomies, and antitheses to enhance reasoning (Danielson & Sinatra, 2017; Grossnickle et al., 2016) .

Also, elaborative feedback may be more important in drawing inferences, or applying of rules in more complex situations (Bangert-Drowns et al., 1991). In the present study, participants completed rather abstract tasks in a short, on-off intervention at the university. For future research, more complex tasks which deal with (simulated) real-life situations are interesting to use to find effect of feedback type on learning, for example in-class debates or small group discussions with immediate process feedback by the teacher. Also, it is unclear whether learning effects persist in the longer term. However, this is an interesting question for future research to address because a central educational goal is to achieve learning effects in the longer term.

The tasks used in our study are used often in confirmation bias studies. However, measuring proneness to confirmation bias with a self-report scale is suggested by Rassin (2008). He states that "the concept of confirmation is multifaceted and therefore by definition not internally reliable" (Rassin, 2008, p. 92). Data from the Rassin's Confirmation Inventory (CI) self-report might be a useful to collect in addition to performance measurements in future research to interpret pretest posttest performance on the four-card selection tasks and hypothesis testing tasks that were used in the current study. However, this measurement needs more validation because the psychometric data of the CI are limited.

Finally, the low Cronbach's α on the posttest and transfer test will be discussed here. Although we used confirmation bias tasks and transfer tasks for the posttest that were validated by previous research (Jonas et al., 2001; Stanovich et al., 2016; Wason, 1968), combining two tasks in the pretest and posttest did not lead to one unidimensional construct that measured the level of confirmation bias.

Retrospectively, this was most probably due to the low level of relevant prior task knowledge, resulting in response guessing and consequently a high degree of non-systematic variance in the pretest sum scores. This was mainly - and unexpectedly - due to the hypothesis testing tasks. However, the other four learning tasks did provide a reliable sum score on the posttest and on these tasks exactly the same effects were found as with the less reliable pre-registered measure. However, for future research, the type of test tasks needs rethinking.

The construct problem also occurred in the transfer test. This might be a result of a priori assuming that these tasks measured a more heterogeneous construct of transferring COS to other biases than the confirmation bias. However, the unreliability is caused by the "Linda"-task, the base rate task and the "extra vert" task by Snyder and Swann (1978).

Significance of the study. Our findings contribute to the existing body of knowledge on teaching CT, especially on instructional design studies including feedback. Feedback was not considered before in confirmation bias research. The results of the present study showed that adding feedback to confirmation bias practice tasks reduces confirmation bias more than practising only but that elaborate feedback in the form of a worked example does not enhance performance more than providing correct answer feedback on practice tasks. This is partly in line with other findings in the feedback literature that show mixed results of various types of feedback and performance outcomes (e.g., Butler et al., 2013; Hattie, 2009; Hattie & Timperley, 2007; Roelle et al., 2017; Shute, 2008). Therefore, further research can focus on increasing the effects of feedback on reducing the confirmation bias through COS.

Moreover, the results from our study are relevant from a methodological perspective. In the vast majority of educational psychology studies, null hypothesis significance testing (NHST), or p-hacking, is used to make inferences about population parameters based on sample statistics. In NHST, a p-value is calculated, which is a conditional probability of observing a particular value of a test-statistic or a more extreme value given the null hypothesis is true and given that the researchers adhere to a predefined sampling and analysis plan. However, researchers often face considerable degrees of freedom when sampling observations and analysing their data. When these researchers' degrees of freedom are not exposed and not taking into account, this will lead to a system in which the probability of false positives becomes very high (see e.g., Nosek, 2017; Simmons et al., 2011). To prevent this inflation of the false positive rate, researchers should pre-register their hypotheses, their sampling plan and their analysis plan. In addition, pre-registration enhances transparency and accountability (Munafò et al., 2017; Nosek & Lakens, 2014; Simmons et al., 2011). Also, our pre-registration identified a problem concerning the reliability of the measurements used. Although the tasks that were used are regarded as typical confirmation bias tasks and reasoning tasks, we were not able to compose a reliable set of items. Based on the pre-registration, we could not ignore this problem but instead this calls for further methodological action to improve these measurements. Without pre-registering research plans, measurement problems will continue to be underexposed within the research community and p-hacking will continue to exist (Chambers & Munaf, 2013; Munafò et al., 2017; Nosek & Lakens, 2014; Simmons et al., 2011). Hence, pre-registration is very helpful in identifying measurement problems in the field. This is an important first step in addressing them. When addressing these issues, data explorations, which are perfectly well possible in pre-registered study, might point at fruitful directions that could be tested in future research. Hence, the present study could serve as an example for other studies because by doing so, we protected ourselves from biasing the results. Pre-registering is one way to actively contribute to ways to achieve more transparency and objectivity in science and in our view, it would be good, if pre-registration would become the norm for educational psychology studies in which hypotheses are tested and counteract the confirmation bias and publication bias in science.

In addition, the findings of the study may also have implications for educational practice. It is important to help students to learn the CT skills that they need to face the future challenges in their work and life. Therefore, they need to transfer what they have learned in school to real-life situations. The findings of our study demonstrate that watching an instructional video and practise tasks with feedback contributes to reducing confirmation bias. However, it is necessary that future research focusses on supporting students in acquisition of transferring these skills in order to enhance unbiased reasoning and decision making. Therefore, it is recommended that cues are given to support transfer skills.

To sum up, learning critical thinking skills such as reducing confirmation bias, is considered to be very hard, because people are resistant to change and have the tendency to cling to their initial beliefs when contradictory information is presented (Douglas, 2000). More research is warranted to prevent people from becoming more divided, but instead let them realize that they can learn to make better decisions from standing in someone else's shoes.

APPENDIX

Four Card Selection Practice Task with an Example of Elaborative as well as Correct Answer Feedback (Translated from Dutch)

Number and colours

On the cards below, there is a number on one face and a colour on the opposite face. Which card(s) do you have to turn over to test the following rule:

"If the card shows an even number on one face, then its opposite face is red."

RED	**YELLOW**	**3**	**8**

Elaborative feedback:

The rule was: "If the card shows an even number on one face, then its opposite face is red". You apply "consider the opposite" by looking for a card that confirms the rule and a card that disconfirms the rule ("the opposite").
In this case, you turn "8" over, because you have to check whether the opposite face is indeed "red" and not accidentally "yellow".
However, you also have to turn "yellow" over because the only way in which the rule can be disconfirmed is to find a card with "8" and "yellow".
"3" and "red" can be ignored, because it has not been said that there should be an even number on a red card and for odd numbers such as "3" there is no requirement at all.

Correct answer feedback:

The correct answer is the "yellow" card, and the "8" card.

Chapter 4

Teaching on video as an instructional strategy to reduce confirmation bias

This chapter has been published as:

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2021). Teaching on video as an instructional strategy to reduce confirmation bias - a pre-registered study. *Instructional Science*, 49(4), 475–496. https://doi.org/10.1007/s11251-021-09547-4

Abstract

The aim of this experiment was to examine the effect of different learning strategies in combination with "consider-the-opposite" instructions on student teachers' confirmation bias. Dutch student teachers (n = 141) took a confirmation-bias pretest and were then randomly assigned to three conditions; teaching on video (TOV), preparing to teach (PTT) and re-study (CC). Students in the TOV and PTT conditions prepared a lesson on "consider-the-opposite" and in the TOV condition they taught this lesson on video. The quality of the prepared lesson was highest for TOV participants suggesting they had gained better understanding of confirmation bias and "consider-the-opposite" than participants in the PTT condition. In addition, the PTT condition scored higher on experienced social presence, whereas the PTT condition and the TOV condition did not differ in self-reported arousal. After the learning phase, all participants took a confirmation bias-posttest and a transfer test. The results showed that confirmation bias was reduced to a similar extent in all conditions. Furthermore, transfer scores were low and did not differ between conditions. We discuss theoretical explanations for the findings from the present study.

Keywords: Confirmation bias, explaining, teaching, video, social presence, arousal, learning, transfer

Introduction

When making decisions and judgments, primary school teachers, like all people, rely on mental shortcuts that facilitate effortless reasoning and usually lead to a suitable response. However, sometimes these cognitive heuristics cause suboptimal reasoning, which might result in cognitive biases, such as the confirmation bias (e.g., Gilovich et al., 2002; Lord et al., 1984; Stanovich, 2011; Tversky & Kahneman, 1974). The confirmation bias refers to the selectivity in finding and using evidence that fits one's own beliefs or hypotheses while neglecting to gather or discard evidence that is opposite to one's own beliefs or hypotheses (Nickerson, 1998; Schwind et al., 2012).

Confirmation bias might have a negative influence on teachers' classroom practice. First, teachers may hold implicit and partly erroneous prior beliefs about certain children, based on for instance socio-economic background or intelligence. Subsequently, these beliefs might guide teachers' reasoning about and behaviour towards these pupils. This in turn might lead to a self-fulfilling prophecy in which children's learning and motivation are negatively affected because they become aligned with the teacher's erroneous prior beliefs (Darley & Gross, 1983; De Boer et al., 2010; Foster et al., 1976; Hornstra et al., 2018; Rosenthal & Jacobson, 1968; Timmermans et al., 2015; Van den Bergh et al., 2010). In addition, the confirmation bias may result in one-sidedness in courses where perspective taking is an important element, such as in civics education (e.g., Abrami et al., 2015; Nickerson, 1998; Schwind et al., 2012).

Perspective taking or "the cognitive capacity to consider the world from another individual's viewpoint" (Galinsky et al., 2008, p. 378), is a critical thinking skill that can be used to mitigate the confirmation bias. It is therefore an important aspect of critical thinking that has to be addressed during teacher education to reduce the confirmation bias and increase teachers' decision making and judgment skills. A general and effective strategy to enhance perspective taking is to consider opposites and alternatives regarding one's initial judgment (e.g., Adame, 2016; Hirt & Markman, 1995; Lord et al., 1984; Lord et al., 1979; Schwind et al., 2012). This typically happens with small interventions consisting of external prompts that stimulate participants to ask themselves when facing a judgment or decision: "What are some reasons that my initial judgment might be wrong?" (Larrick, 2004, p. 323). Although studies with these consider-the-opposite interventions (e.g., Hirt & Markman, 1995; Lord et al., 1984) have provided important theoretical insights into the mechanism underlying the confirmation bias, they are of limited use for educational practice.

This is because in real-life settings, people have to learn to prevent a confirmation bias without receiving an external prompt. Therefore, studies with an educational focus are aimed at uncovering instructional design principles that help people to reason unbiasedly in the absence of such prompts. These studies are typically knowledge-rich in the sense that participants are provided with background knowledge about the confirmation bias (e.g., what is it, and why is it important to reduce the confirmation bias) and subsequently receive an unbiased reasoning instruction in combination with practice. Previous studies on bias reduction in general (e.g., Adame, 2016; Dunbar et al., 2014; Heijltjes, Van Gog, Leppink, et al., 2014; Mussweiler et al., 2000; Sellier et al., 2019) and on confirmation bias in particular (Jonas et al., 2001; Lord et al., 1984; Lord et al., 1979; Van Brussel et al., 2020) have shown that a debiasing effect occurs when providing participants with (1) general instructions on the importance of unbiased reasoning and with worked examples, (2) the possibility of practicing with relevant unbiased reasoning tasks and (3) feedback on the practice tasks.

Participants in instructional design studies on reducing confirmation bias were allowed to decide for themselves how they processed the instruction (e.g., Adame, 2016; Lord et al., 1984; Van Brussel et al., 2020). It has been suggested though, that learners who are given this freedom of choice often rely on popular but suboptimal learning strategies that lead to passive or superficial learning (e.g., Blasiman et al., 2017; Dirkx et al., 2019; Karpicke et al., 2009; McCabe, 2011; Morehead et al., 2016). Renkl (1997, 2002) for example, has shown that, when left to their own devices, only a small amount of students deployed high quality self-explanations of worked examples (e.g., more principle-based explanations about probability calculation) that support learning more compared to the largest group that failed to make effective use of worked examples.

Based on the considerations above, it might be expected that adding a generative learning activity to a “consider-the-opposite” instruction might lead to better performance on confirmation bias tasks. When learning is regarded as a generative activity, it “involves actively constructing meaning from to-be-learned information by mentally reorganizing it and integrating it with one’s existing knowledge” (Fiorella & Mayer, 2016, p. 717). This implies that learning not only depends on how information is presented to learners, but also on how a learner can transform incoming information into useable knowledge (Fiorella & Mayer, 2016). Generative learning strategies that include generating explanations and explaining to others are effective for knowledge building and have more robust effects on memory and transfer of knowledge compared to more superficial learning strategies, such as (massed) re-study (e.g., Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019; Kobayashi, 2019; Renkl, 1997; Roscoe & Chi, 2008).

Therefore, to enhance the reduction of confirmation bias, it is important to encourage students to use effective learning strategies (Dunlosky et al., 2013). For example, the learning strategy "learning by teaching" engages a learner in the learning material and is assumed to be effective for building meaningful knowledge structures (see also Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019; Kobayashi, 2019; Roscoe & Chi, 2008). Learning by teaching can be described as teaching to-be learned material with the goal of helping others learn (Fiorella & Mayer, 2016). However, the "teacher" also benefits from explaining to others because one has to select the relevant information to include in the explanation, organizing it in a way that it can be understood by others, and one has to elaborate on the material by incorporating one's existing knowledge, which leads to new knowledge structures (Duran, 2017; Fiorella & Mayer, 2016).

Learning by teaching on video. One fairly new and promising learning by teaching strategy is learning by teaching on video. This strategy has shown promising effects on learning and transfer across various ages, different learning materials and in various domains compared to summarizing or restudying the learning material (Fiorella & Mayer, 2013; Hoogerheide et al., 2016; Hoogerheide et al., 2014a; Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019).

Learning by teaching on video consists of two learning phases: The preparation of the lesson and the actual act of teaching to a (fictitious) peer audience. In the preparation phase, learners need to self-explain and e.g., generate analogies to make abstract concepts more apprehensible, or use examples to illustrate the to-be taught concepts which supports knowledge building and learning (Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Hoogerheide et al., 2016; Roscoe & Chi, 2008). During the preparation phase, one has to monitor one's own knowledge and understanding through the retrieval of prior knowledge to support one's own knowledge building. Prior knowledge must be integrated with the new knowledge, which might result in detecting knowledge gaps that have to be repaired before the actual act of teaching takes place because the explanation must be accurate and complete (Dunlosky et al., 2013; Fiorella & Mayer, 2013, 2014; Roscoe & Chi, 2008). During the teaching phase, one directs oneself to a camera and teaches the prepared lesson to a fictitious audience. On top of the direct cognitive elements that are evoked by preparing a lesson, the beneficial learning effects of the actual act of teaching on video might be enhanced by indirect elements that are evoked by feelings of social presence by addressing an audience. Feelings of social presence might in turn generate arousal which may result in better learning and transfer (Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019).

Social presence can be described as being aware of the fictitious presence of others in an (virtual) environment (e.g., Biocca & Harms, 2002, October; Gunawardena, 1995; Lowenthal, 2012; Slater et al., 1999). Considering and being aware of one's audience by teaching on camera to a fictitious audience, induces feelings of social presence that might in turn lead to an increased level of arousal (Hoogerheide et al., 2016; Hoogerheide et al., 2014b; Hoogerheide, Renkl, et al., 2019). The physiological construct of arousal is defined by Hoogerheide et al. (2019; Sharot & Phelps, 2004) as a state of being excited or activated with increasing heart rate, blood pressure, and perspiration as a result. Arousal can affect task performance because one is stimulated to generate accurate and complete explanations to ensure that the audience understands the learning material. Researchers have proposed that this leads to a more effective use of working memory capacity, leading to the construction of meaningful knowledge in long-term memory (Hoogerheide, Renkl, et al., 2019). However, although higher levels of arousal and higher posttest performance were observed in the teaching on video condition compared to the restudy control condition, Hoogerheide, Renkl et al. (2019) found no association between arousal and isomorphic and transfer task performance. Therefore, it is still unclear whether or not arousal is beneficial for performance.

Nevertheless, recent studies have shown promising results of teaching on video compared to re-studying (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2016; Hoogerheide et al., 2014b; Hoogerheide, Renkl, et al., 2019). In these studies, teaching on video resulted in higher posttest performance over only the expectancy to teach (i.e., preparing to teach), or re-study in the domain of text-based and example-based learning of physics and electrical engineering subjects. The learning effects of teaching on video were especially beneficial for learners with low prior knowledge (Hoogerheide, Renkl, et al., 2019).

The present study[7]. Returning to the question of how to optimise instructional design to reduce the confirmation bias, it can be hypothesised that this bias might be more strongly reduced if participants are being incited to use effective learning strategies during training, such as learning by teaching on video. In the present study, this issue was addressed with student teachers as participants. For these students, using teaching on video, as compared to a training without teaching on video might be particularly useful as it creates an authentic training situation (i.e., it mimics the task they have to perform during their future job). A part as large as 40% of the Dutch teacher education curriculum takes place during teaching internships within primary schools. Mimicking teaching while learning by teaching on video might therefore be seen as an additional meaningful learning opportunity for student teachers.

Given the positive effects of earlier studies (Adame, 2016; Lord et al., 1984; Van Brussel et al., 2020) on reduction of the confirmation bias by using a direct "consider-the-opposite" instruction, subsequent practice and feedback, and considering the promising effects of teaching on video on learning and transfer, the present study focused on the question whether teaching on video is an effective learning strategy to reduce the confirmation bias. In addition, it was explored whether feelings of social presence and arousal might be associated with possible differences between learning by teaching on video compared to learning through preparing to teach.

To address the research question on learning, participants completed a pretest consisting of confirmation bias tasks. Subsequently they received a "consider-the-opposite" instruction, practice tasks with feedback after which participants were assigned to one of three conditions: teaching on video, preparing to teach, and re-study (control condition). The first and second group prepared a lesson based on generated self-explanations based on the provided instruction and practice material. In addition, participants in the Teaching on video condition taught the prepared lesson to a fictitious audience by recording a video. To explore any differences, feelings of social presence in the Teaching on video and Preparing to teach conditions, and self-reported arousal levels in all three conditions were measured with questionnaires. Participants in the control condition re-studied the instruction and practice tasks with feedback. Afterwards, all participants completed a confirmation bias-posttest to assess learning and a transfer test.

7 The present study was pre-registered at the Open Science Framework. Hypotheses, planned data-collection, planned methods, and planned analyses can be found on https://osf.io/u9w4v/?view_only=09ee01a4038b4b59be454655aca68ab9.

Hypotheses. Based on the considerations above, it was hypothesised that student teachers would learn more effectively to reduce their confirmation bias and transfer their learning through teaching on video (A), compared to preparing to teach (B) because of the social presence hypothesis: the awareness of a fictitious audience during teaching induces motivational processes that support learning and transfer in the teaching on video condition more than when only preparing to teach (Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019). Furthermore, it was hypothesised that student teachers learn more effectively to reduce their confirmation bias and transfer their learning through preparing to teach (B) compared to a control condition in which participants re-study the learning material (C) because of the generative learning hypothesis: student teachers who will learn by preparing to teach benefit from generative processes that are effective for deep learning such as discovering and repairing knowledge gaps, and organising the information into meaningful knowledge structures (Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019). In sum, for learning and transfer it was expected that A > B > C.

Method

Participants and design. Participants (n = 141, M_{age} = 19,9, SD_{age} = 4,59) were student teachers from a Dutch primary education institution at university of applied sciences level. They were randomly assigned to one of the three conditions: teaching on video (n = 46), preparing to teach (n = 44) and re-study (n = 51). To obtain a power of .80 for the interaction between learning strategy and test moment, under a significance level of .05, a correlation of .09 between repeated measures (estimate based on similar experiments by the authors) and a medium effect size (f = .27, based on Hoogerheide et al., 2019), we needed to test at least 138 participants according to G*Power (Faul et al., 2007). Hence, the current sample was sufficiently large to detect a medium effect size on the crucial interaction effect on learning and transfer. All students gave informed consent prior to the experiment.

Materials. Both tests, the instruction and the practice tasks with feedback were based on prior research by the authors and were delivered through the Qualtrics platform (Qualtrics, 2017). In the Teaching on video condition, the actual act of teaching was recorded on participants' smart phones and sent to the researcher via WhatsApp. As promised to the participants, the videos were deleted immediately after being viewed by the researcher. Participants in the teaching on video and preparing to teach condition prepared their lesson by typing the preparation in

Qualtrics. Both groups received an envelope with two text-based tasks that needed to be explained in the preparation of the lesson. Participants in the control condition received an envelope with the text-based instruction and practice tasks with feedback to re-study.

Test and practice tasks. To determine the level of confirmation bias in the pretest and posttest, two task types were used that are commonly known to assess confirmation bias: three hypothesis testing tasks (As used before in e.g., Jonas et al., 2001; Van Brussel et al., 2020) and three Wason's four-card selection tasks (As used in e.g., Sellier et al., 2019; Van Brussel et al., 2020; Wason, 1968) per test.

The transfer task was added to the posttest and was based on the study by Snyder and Swann (1978) and was used before in a "consider-the-opposite" study (Van Brussel et al., 2020).

For practice, one hypothesis testing task and one four-card selection task were used. Feedback in the form of a worked example that was based on the instruction was provided after each practice task in all three conditions. See Appendix A and B for task examples.

For the Teaching on video and the Preparing to teach participants, one hypothesis testing task and one four-card selection task without the answer or feedback were provided on paper that needed to be explained during the intervention.

The hypothesis testing tasks based on Jonas et al. (2001), showed how the biased search for information is related to someone's position on a statement. In a hypothesis testing task, participants were asked whether or not they agreed with a statement (e.g., "Nowadays, children have less perseverance than children from the past") and were asked to motivate their answer to stimulate them to think seriously about their statement. After this, participants read eight short summaries of articles about this subject. Four articles were strongly in favour and four were strongly opposed to the statement. Based on the given summaries, participants chose four articles which they wanted to read to gather more information about the statement. After selecting, the task was finished. Note that participants did not actually receive the articles. Confirmation bias was expressed by participants if they chose information that agreed with their initial point of view. If their choice was more balanced (choosing e.g., two articles in favour and two articles contra), their confirmation bias was mitigated.

For the four-card selection tasks, participants had to choose which of the four given cards they needed to turn over to test a given rule. In general, participants have the tendency to select two affirmative cards, however, in order to test the rule, they need to turn over an affirmative card as well as a non-affirmative card. The tendency to only select the affirmative cards can be seen as an expression of confirmation bias (Nickerson, 1998). For example, the given rule is: "If the card shows an even number on one face, then its opposite is red". The four cards that are shown are "Red", "Yellow", "3", and "8". The only possible way to falsify the rule is by turning the affirmative card ("8" in the example) and by turning a non-affirmative card ("not red" = "yellow" in the example). However, participants usually choose the cards that confirm the rule rather than cards that disconfirm, which is an expression of confirmation bias.

The transfer task differed in structure, but could be solved with "consider-the-opposite" to reduce the confirmation bias (Lord et al., 1984; Snyder & Swann, 1978). Participants were provided with information about extravert behaviour. They had to choose 12 out of the 26 provided questions (11 on extraversion, 10 on introversion and five neutral questions) to interview a person to find out whether this person is an extravert. Confirmation bias might occur here, because people have the tendency to choose questions that confirm the extraversion (e.g., "What would you do to enliven a boring party?") and avoid asking questions that must prove the opposite (e.g., "Which things do you dislike about a noisy party?").

All participants could score a maximum of 9 points for the pretest and posttest, and two points the transfer posttest. For each correct answer on an isomorphic Wason's four-card selection task participants received one point. The answer was correct when participants chose one card which confirmed the rule and one that disconfirmed the rule. For the isomorphic hypothesis testing tasks based on Jonas et al. (2001), participants received no point, one or two points, depending on the information they selected: participants who selected four pieces of information that were either all pro or contra the statement, received no points. If they selected two pieces of information in favour and two pieces of information that disagreed with their statement, participants received two points because that showed a tendency for perspective taking. If they selected three pieces of information in favour and one not in favour of their initial statement, they earned one point because that means they still showed too much confirmation bias. For the transfer task, participants earned no points if they chose nine or more than nine questions that focused on one side of the topic, two points if they choose between four to eight questions that focused on the topic and one point if they chose three or less focused questions.

Instruction. In the text-based instruction, confirmation bias was defined, and it was explained what the importance is of reducing confirmation bias, and that the critical thinking skill perspective taking supports this. Furthermore, it was explained how confirmation bias emerges in the information search process related to a hypothesis testing example from the pretest. Next, a step-by-step explanation of "consider-the-opposite" as a strategy to reduce confirmation bias was given. The first step was to ask oneself the question: "What are possible reasons that my initial judgment might be wrong?". Secondly, one had to take perspective by considering the opposite which means that one must consider contradictive information. Finally, one has to critically assess the contradictive information and weigh both perspectives before making a final decision.

Social presence and arousal. Questionnaires were used to explore whether feelings of social presence and arousal levels might explain the possible beneficial effects of Teaching on video, and differences between Teaching on video compared to Preparing to teach or Re-study.

Feelings of social presence were measured with a questionnaire in the Teaching on video condition and Preparing to teach condition (Cronbach's alpha = .73). Participants had to indicate the extent of their level of feelings of social presence. Based on the factors of social presence mentioned in the overview by Cui (2013), participants indicated on a scale from 1 (strongly agree) to 5 (strongly disagree) to which degree the statements represent how they felt after the learning phase. See Appendix C for the statements.

To measure arousal, the activation-deactivation adjective check list (ADACL, see Appendix D) by Thayer (1967, 1986) was used in all three conditions. This checklist was translated to Dutch and assesses core arousal or activation states based on two dimensions (activation and deactivation) and four factors. For general activation, energy was measured with adjectives. For deactivation-sleep, adjectives regarding tiredness were measured. High activation was measured by tension adjectives and general deactivation by calmness adjectives. Previous studies revealed Cronbach's alphas of .92 on "energy", .89 on "tension" and "calmness", and .92 on "tiredness" (Boyle et al., 2015; Thayer, 1978). Participants rated on a four-point scale how well the adjective described their immediate feelings after the learning phase (Teaching on video, or Preparing to teach, or Re-studying). The scale consists of the terms "definitely feel", "feel slightly", "cannot decide", and "definitely do not feel this way". The ADACL was used to investigate whether arousal level differences between the conditions could explain different effects of the instructional strategies.

The validated ADACL by Thayer (1978, 1986) was scored by assigning points to the following response categories: four to “definitely feel”, three to “feel slightly”, two to “cannot decide”, and one to “definitely do not feel” and averaging the five scores of each subscale. “Wakeful” and “wide-awake” were reversed for the Tiredness subscale.

Procedure. Prior to the experiment, participants were randomly assigned to one of the three conditions: Teaching on video (A), Preparing to teach (B), or the control condition in which participants re-studied the learning material (C). The experiment was conducted at a Dutch primary education institution. At the time of the experiment, “consider-the-opposite” to reduce confirmation bias was not yet taught in the curriculum. The student teachers were instructed that this experiment was conducted for educational research purposes to improve the curriculum, that the results are not used for performance evaluations and that anonymity is guaranteed. Participants in the Preparing to teach and Re-study condition were tested group wise in a classroom. Participants in the Teaching on video were seated in separate soundproof cubicles because they had to record their explanation.

At the start, the experimenter provided a short, general verbal instruction about the nature of the experiment and explained some rules (e.g., it is not permitted to open websites other than Qualtrics and Word, or to leave the room during the experiment). Participants were asked to open the link that led to the Qualtrics website. The experiment started with six tasks of the pretest. After this, participants read the instructional text. Next, participants completed the two practice tasks and read the worked example feedback after each task. Then the three conditions each followed their own intervention: Teaching on video, or Preparing to teach, or Re-study. The total time given for the intervention phase was ten minutes. In the Preparing to teach and Re-study group time was controlled for by the Qualtrics platform. Teaching on video participants were asked to set eight minutes on their smart phone to prepare the lesson in Qualtrics and afterward to set two minutes on their smart phone to teach the lesson on video. This was necessary because it was technologically not possible to divide the learning phase for this group in eight and two minutes.

In the Teaching on video group, participants received the instruction to prepare and teach a lesson for peers, based on the “consider-the-opposite” instruction and the practice tasks with feedback. They also received a hypothesis testing task and a four-card selection task that they needed to explain to their peers (see Appendix A and B for task examples). The preparation was typed in Qualtrics and the teaching was recorded with the camera of the participants’ own smart phone. In the Preparing to teach group, participants received the same instruction as in the Teaching on video but without the teaching assignment. Participants in the Teaching on video group received eight minutes to prepare and two minutes to teach. Participants in the Preparing to teach group were given ten minutes for their preparation. Participants in the Re-study control condition were given ten minutes to study the learning material (instruction and practice tasks with feedback). Participants in the control group indicated afterwards how many times they read the information.

Finally, participants in the Teaching on video and Preparing to teach condition completed the social presence test. All participants took the ADACL arousal questionnaire and posttest, and were asked for their prior education, sex and age. Participants were thanked and given the opportunity to receive the results of the research.

Data analysis[8].

Analysis of the preparation of the lesson. The participants in the Teaching on video (n = 46) and Preparing to teach condition (n = 44) prepared a lesson and the teaching group used this preparation to record the lesson. An independent rater and the first author scored 22% of the texts to check for the reliability of the scoring. Results showed the interrater reliability was substantial for "completeness" ($\kappa = 0.74$) and moderate for "accuracy" ($\kappa = 0.55$) and therefore one rater scored the remainder of the preparation texts.

The approach for the qualitative analysis of the preparation texts was based on Hoogerheide, Renkl, et al. (2019). The 90 written products were scored on the degree of presence of four elements from the instruction and practice phase: the definition of confirmation bias and the three elements of the "consider-the-opposite". The explanation of the example tasks was scored on "completeness" and "accuracy". For each preparation, this means that "completeness" was scored as how many of the total of three steps from "consider-the-opposite" were counted in the preparation (no missing elements = 3 points) and whether the two examples were explained (no missing elements = 2 points) resulting in a range from 0 to 5 for completeness. "Accuracy" focused on the quality of the explanation: no mistakes in the explanation (1 point), the to-be explained elements were partially accurate (one mistake; 0.5 points), or inaccurate (multiple mistakes; 0 points). Both measures were converted

8 The assumptions on which the choices for our data analyses are based, can be found in the pre-registered manuscript at the Open Science Framework: https://osf.io/u9w4v/?view_only=09ee01a4038b4b59be454655aca68ab9

to percentages to be able to compare both scores (i.e., multiply "completeness" scores by 20 (5x20= 100) and "accuracy" scores by 100 which results in a score of 0, 50 or 100%. To determine whether there were differences between the teaching and preparation groups on the quality of the preparation, One-Way ANOVAs were conducted.

Pretest and posttest analysis. We expected low Cronbach's alpha on the pretest due to low relevant prior task knowledge, based on the pretest scores with similar tasks of a previous conducted study (Van Brussel et al., 2020). Indeed, the pretest had a low Cronbach's α of .51. Therefore, we calculated the total mean score of the pretest and used it as baseline score of the relevant prior knowledge level. To allow for a comparison between the pretest baseline and the posttest score, we multiplied the pretest baseline score with 7/9. This was done, because 1 of the six posttest items was removed from the analysis. As a result of this removal, the total points on the posttest were 7/9 times the number of total points on the pretest. After the correction, the pretest baseline score was 3.17.

For the posttest, the minimal lower limit for reliability was set on a Cronbach's α of .60. Our analysis however, yielded an unreliable item set. One item showed a negative item-total correlation. This was a hypothesis testing task with the statement "Biotechnology can solve the food problem in Africa". We deleted this item and the remaining subset of 5 items had a Cronbach's α of .61. Our analyses were conducted on this subset.

Questionnaires analyses. Items 4, 9 and 10 from the social presence questionnaire (for the Teaching on video and Preparing to teach groups only), were pooled before analysis. The lowest average per item was 1, and the highest was 5 which means that participants could score between 10 and 50 points. The higher the score, the lower the feelings of social presence. For each participant, an average score was calculated by dividing the total score by 10.

Results

All participants (n = 141, 112 women) completed the experiment and gave informed consent at the start. The mean time spent was 59 min (SD = 13.11) in which the intervention phase for all three groups was set at 10 min. In all analyses below, a significance level of .05 was used. Partial eta-squared (η_p^2) is reported as measure of effect size for the ANOVAs for which .01 is considered small, .06 medium, and .14 large.

See Table 1 for means and standard deviations of the posttest, transfer test, social presence questionnaire and ADACL questionnaire per condition.

Table 1. Pretest Baseline Score, Mean Scores and Standard Deviations of the posttest, Corrected Posttest, Transfer Test, Social Presence Questionnaire and ADACL Questionnaire per Condition.

	Teaching on Video	Preparing to Teach	Re-study
	n = 46	*n* = 44	*n* = 51
	M (SD)	*M (SD)*	*M (SD)*
Pretest baseline score	3.17 (1.86)	3.17 (1.86)	3.17 (1.86)
Posttest (based on 5 items)	4.59 (1.51)	4.95 (1.74)	4.78 (1.47)
Transfer-test (1 item)	1.63 (0.77)	1.77 (0.64)	1.69 (0.74)
Social Presence (10 items)	2.49 (0.66)	2.76 (0.51)	
ADACL Energetic (5 items)	2.83 (0.88)	2.55 (0.79)	2.33 (0.81)
ADACL Tiredness (5 items)	0.89 (0.92)	1.30 (0.93)	1.45 (0.83)
ADACL Tension (5 items)	2.02 (0.88)	1.80 (0.70)	1.55 (0.58)
ADACL Calmness (5 items)	1.89 (0.71)	2.05 (0.65)	2.45 (0.64)

The completeness and accuracy of the preparation of the lesson. The 90 participants of both Teaching on video and Preparing to teach prepared a lesson after the learning phase. The mean wordcount of the 46 texts of the Teaching on video participants was 173 words, compared to 166 words of the 44 texts in the Preparing to teach condition. Note that the latter group was able to spend ten minutes on writing the preparation compared to eight minutes for those who also had to teach the lesson. A One-Way ANOVA showed that there was no significant difference between conditions on "completeness" $F(1,88) = 3.87$, $p = .052$, with a large effect size of $\eta_p^2 = .14$ and a mean score for the Teaching on video group of 68% ($M_e = 60$, $SD = 24.79$) and 45% % ($M_e = 60$, $SD = 28.15$) for the Preparing to teach group. On "accuracy" Teaching on video participants significantly outperformed Preparing to teach participants $F(1,88) =$,8.48 $p < .05$, with a small effect size of $\eta_p^2 = .01$. The mean score for the Teaching on video group was 45% ($M_e = 50$, $SD = 41.13$) compared to 22% ($M_e = 0$, $SD = 33.06$) for the Preparing to teach group.

Feelings of social presence. A One-Way ANOVA showed that the difference between these two groups on feelings of social presence was significant: $F(1, 88) = 4.69$, $p < .05$ with a small effect size of $\eta_p^2 = .05$. In contrast to what was expected, participants in the Preparing to teach group reported higher feelings of social presence compared to the Teaching on Video participants (see Table 1).

Arousal. A One-Way ANOVA was conducted to explore the effect of Condition on arousal, as measured by the ADACL subscales (Thayer, 1986). There was a statistically significant difference on the four subscales of the ADACL for the three conditions: Energetic subscale: $F(2,138) = 4.28$, $p < 0.5$ with a medium effect size of $\eta_p^2 = .06$. Tiredness: $F(2,138) = 4.97$, $p < .05$ again with a medium effect size of $\eta_p^2 = .07$, Tension: $F(2,138) = 5,14$, $p < .05$ with a medium effect size of $\eta_p^2 = 07$ and Calmness: $F(1, 138) = 9.24$, $p < 0.5$ with a medium effect size of $\eta_p^2 = 0.12$.

Post-hoc comparisons using Tukey HSD tests for each ADACL subscale revealed that participants in the Teaching on video and Preparing to teach groups reported higher energetic and tension levels and lower tiredness and calmness levels compared to the control condition. On all four subscales, Teaching on video participants scored significantly higher on energy and tension and lower on tiredness and calmness compared to the Re-study control condition. On the calmness subscale we also found that participants in the Preparing to teach group scored significantly lower compared to the Re-study group. No significant difference was found between Teaching on Video and Preparing to Teach on all four subscales.

Learning. A paired samples t-test showed that the mean score on the five items of the posttest, collapsed across conditions, ($M = 4.77$, $SD = 1.57$) was higher than the corrected baseline score of the pretest (3.17). This difference was significant, $t(140) = -12.30$, $p < .05$, with a small effect size of Cohen's $d = .23$. Contrary to what was hypothesised, the analysis with the five posttest items showed no significant difference between the three conditions: $F(2,138) = .616$, $p = 0.542$ with a small effect size of $\eta_p^2 = .01$ which means that we have no evidence for a difference between the three conditions on the confirmation bias tasks of the posttest.

Transfer. A One-Way ANOVA on the transfer task showed that there was no significant difference between the three conditions on transfer: $F(2, 138) = .446$, $p = .641$ with a small effect size of $\eta_p^2 = .01$.

Conclusion and discussion

The aim of this pre-registered experiment was to examine the effect of learning by teaching on video on reducing the confirmation bias in student teachers. By pre-registering and storing the data on the Open Science Framework, we protected ourselves from biasing the results by e.g., null hypothesis significance testing or p-hacking. Furthermore, we shared our data on the Open Science Framework. We did so because preregistration and open science are important ways to actively contribute to ways to achieve more transparency and objectivity in science and in our view; it would be good if it became the norm for educational psychology studies to counteract the confirmation and publication bias (Munafò et al., 2017; Nosek, 2017; Nosek & Lakens, 2014; Simmons et al., 2011).

Based on the social presence hypothesis and the generative learning hypothesis, we expected that participants who learned through teaching to a fictitious audience on video would perform better on learning and transfer task on a confirmation bias-posttest compared to participants who only prepared a lesson, or participants who restudied the instruction. With respect to learning, we found that collapsed across conditions, participants were less susceptible to confirmation bias on the posttest than on the pretest. This finding is in line with earlier instructional “consider-the-opposite” studies (Adame, 2016; Hirt & Markman, 1995; Lord et al., 1984; Mussweiler et al., 2000; Van Brussel et al., 2020) . Furthermore, to prepare a lesson, the Teaching on video group and the Preparing to teach group had to retrieve relevant information from the instruction and practice, information about what the confirmation bias is and how to prevent it through considering the opposite, and apply it to new cases. The accuracy of the lesson preparations was higher in the Teaching on video condition than in the Preparing to teach condition. This finding might suggest that Teaching on video is more beneficial for the build-up of conceptual knowledge of confirmation bias and considering the opposite than preparing to teach. However, it might also be possible that the accuracy differences we found between to Teaching on video and the Preparing to teach condition reflect differences in reporting. Perhaps, participants in the Teaching on video condition checked their explanations more extensively – which would lead to higher accuracy scores - than their peers in de Preparing to teach condition because they had to report to a fictitious audience. In any case, it would be interesting to further investigate whether teaching on video and preparing to teach have differential direct or indirect effects on the construction of conceptual knowledge.

Although all conditions showed less confirmation bias on the posttest than on the pretest, the three conditions did not differ on their mean posttest scores. This was contrary to our predictions, as we had expected for learning and transfer that the Teaching on video condition would perform best, followed by the Preparing to teach and the restudy control condition. There are several explanations for these unexpected findings. First, the lack of difference between the two generative conditions, i.e., the Teaching on video condition and the Preparing to tech condition, and the restudy condition might the retrieval hurdle that participants had to pass in former conditions (Roelle & Nückles, 2019). Although a plethora of research findings has demonstrated that generative learning strategies lead to better learning outcomes than repeated study, under certain conditions, this positive effect might be reduced or erased when generative learning strategies are combined with retrieval practice. Specifically, when mental representation of the content after instruction and practice is relatively low, then engaging in retrieval practice and generative activities is not more effective than restudying because learners cannot retrieval all required information from memory and/or because retrieving information takes so much time that it prevents participants from spending sufficient time on generative activities. Perhaps, this occurred in the Teaching on Video condition and the Preparing to teach condition in the present study, which involved retrieval practice in combination with a generative learning activity. In contrast, in studies that did find beneficial effects on learning or transfer by teaching on video, participants were allowed to use the instructional material for their preparation (e.g., Hoogerheide, Renkl, et al., 2019; Roscoe, 2014; Roscoe & Chi, 2008). Thus, in the present study, the retrieval hurdle might have suppressed the beneficial effects on learning and transfer as hypothesised for Teaching on video and Preparing to teach compared to Re-study.

A second explanation for the absence of significant learning and transfer performance between the three groups might lie in the types of tasks on the final test. Although these tasks are generally used to measure reasoning and confirmation bias, research often fails to find differences between groups on learning strategies that are effective in other learning contexts and content (e.g., Heijltjes, Van Gog, Leppink, et al., 2014; Jonas et al., 2001; Van Brussel et al., 2020; Van Peppen et al., 2018). For the hypothesis testing tasks as well as the Wason four-card selection task, all participants might have developed heuristics during the intervention phase. If these heuristics were strong enough to produce relatively high levels of success during the final test, then this might have considerably reduced any additional effect of explaining new tasks in the Teaching on video and Preparing to teach conditions. In combination with the reliability challenges of the pretest and posttest (Van Brussel et al., 2020), it is suggested that for future research alternative tasks and measurements of confirmation bias and perspective taking are used. Fiorella and Mayer (2013, 2014), and Hoogerheide and colleagues (2016; 2014b; 2019; 2019) used conceptual knowledge tasks to measure learning and participants studied a text or worked examples of scientific topics before being assigned to either teaching conditions, preparing conditions, summarizing conditions or re-study conditions. Using more applied tasks has not yet been considered in learning by teaching research. In addition, solving more applied tasks based on real-life scenarios in combination with a teaching a real audience might enhance understanding when to apply "consider-the-opposite" to take perspective when confirmation bias is at risk (based on the meta-analysis on critical thinking instruction by Abrami et al., 2015).

A third explanation of the absence of differences on learning and transfer, might the timing of the explaining activity. In our study, participants explained by preparing and teaching after the learning phase. A recent study by Lachner et al. (2019), however, suggested that the timing of the explaining activity matters with regard to performance. This research showed that learning is enhanced if students explain in between learning activities instead of afterwards. Although Lachner et al. (2019) did not use learning by teaching on video, the cognitive processes that are supported by moving the explaining to an earlier learning phase might also lead to a beneficial effect with teaching on video since the timing of explaining plays an important role on learning. For future research, the timing of explaining might be considered.

The present study was also set up to examine the social presence hypothesis, which suggests that the awareness of a fictitious audience during teaching can improve learning because of an increase in arousal levels and subsequent beneficial cognitive processes (Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019) However, measuring social presence directly and arousal with the ADACL was not done before in learning by teaching research. The outcome of the self-reported feelings of social presence, arousal questionnaire and learning outcomes of our study do, however, not provide evidence for the social presence hypothesis; social presence scores were highest in the Preparing to teach group instead of the Teaching on video group, arousal levels did not differ between the three groups and no significant differences on learning and transfer were found.

In our study as well as in the teaching on video studies by Hoogerheide et al. (2016), Hoogerheide, Renkl et al. (2019), and Hoogerheide, Visee et al. (2019), participants did not address a real audience. We might therefore question how strong the learning effect of feelings of social presence and arousal is when participants teach to a fictitious audience. Hence, a suggestion for future research is to increase the differences between the two strategies (e.g., instruction and practice in a real-life setting with a present audience versus preparing to teach only, or self-explaining) to further investigate the role of social presence and arousal in learning by teaching. For future research, it is of importance to replicate our social presence and arousal measurements in order to determine the robustness of our findings (Lakens et al., 2012; Schmidt, 2009; Simmons et al., 2011).

In sum, our findings point to the direction that learning by teaching on video enhances understanding of the learning material but in the current study did not lead to better performance on confirmation bias tasks compared to only preparing to teach or re-study. Our results did not support the social presence hypothesis because feelings of social presence and arousal did not differ between participants who prepared and taught compared to those who only prepared. Moreover, and in contrast to other recent learning by teaching on video studies (e.g., Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019), social presence and arousal were measured through questionnaires. More research is needed to test the robustness of these measurements. In addition, it is interesting to further investigate the mechanisms responsible for possible benefits of learning by teaching on video for student teachers with more applied tasks to learn how to reduce the confirmation bias. This is important, because the confirmation bias can have detrimental effects for children the classroom due to their teachers' biased judgment and decision making.

APPENDIX A

Example of a Four Card Selection Task (Translated from Dutch)

Number and colours

On the cards below, there is a number on one face and a colour on the opposite face. Which card(s) do you have to turn over to test the following rule:

“If the card shows an even number on one face, then its opposite face is red”

RED	**YELLOW**	**3**	**8**

APPENDIX B

Example of a Hypothesis Testing Task (Translated from Dutch)

“Learning styles”
You are a student teacher, and you have your internship in group 6 of a teacher called Benjamin. He is your mentor this year. Benjamin wants to make the students more independent so that the transition to the secondary school will be smoother. He therefore wants to visualise the learning styles of the children. His vision is that responding to learning styles of children helps them to learn.

He asks you as a student to look for theory on this subject so that he can inform himself.

Below is a summary of the information from eight sources that report on “learning styles”. Tick which four sources you choose to share with Benjamin.

Example of two sources of which participants could choose from:
- In a 2008 analysis, four psychologists have assessed every study carried out since 1920 on learning styles. The psychologists therefore conclude that “at present there is insufficient scientific basis to effectively use learning styles in general teaching practice. The limited resources could be better used for educational practices that have a strong scientific basis”. In short, this research does not support the vision of your mentor.

- In a scientific research into opinions and the behaviour of students with regard to their learning, it appears that there is a correlation between the motivation and the learning activities of students. This coherence is such that it can speak of four learning styles: the reproductive learning style, the meaning-oriented learning style, the application-oriented learning style and the unfocused learning style. In short, this research supports the vision of your mentor.

APPENDIX C

Social Presence Questionnaire (Translated from Dutch)

Please indicate to which degree the following statements represent how you felt during the learning phase

(1) = strongly disagree
(2) = disagree
(3) = undecided
(4) = agree
(5) = strongly agree

1. I was aware of an audience / social presence during the explaining assignment [+]

2. I thought about what would happen if I did not understand the learning material [+]

3. I was motivated to accurately explain the learning material to my peers [+]

4. I did not feel engaged with my peer audience while explaining [-]

5. I was motivated to make a complete explanation for my peers [+]

6. I explained my lesson explicitly with my peer audience in mind [+]

7. My peer audience stimulated me to do the best I can [+]

8. I felt some tension while explaining because the learning material was rather complex [+]

9. That a peer audience will watch my explanation, does not bother me at all [-]

10. Teaching my peer audience, hampered my own learning [-]

APPENDIX D

Activation Deactivation Adjective Check List (Thayer, 1986)

Each of the words here describe feelings or mood. Please use the rating scale next to each word to describe your feelings at this moment. Work rapidly, but please mark all the words. Your first reaction is best. This should take only a minute or two.

(1) = definitely feel (2) = feel slightly (3) = cannot decide (4) = definitely do not feel

1. Active
2. Placid
3. Sleepy
4. Jittery
5. Energetic
6. Intense
7. Calm
8. Tired
9. Vigorous
10. At-rest
11. Drowsy
12. Fearful
13. Lively
14. Still
15. Wide awake
16. Clutched up
17. Quiet
18. Full of pep
19. Tense
20. Wakeful

Chapter 5

Comparing instructional strategies to support student teachers' learning to prepare an open-minded citizenship education lesson

This chapter has been resubmitted after revision as:

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (submitted). Comparing instructional strategies to support student teachers' learning to prepare an open-minded citizenship education lesson.

Abstract

Open-mindedness is defined as one's willingness and ability to consider opposing beliefs and perspectives and give them a serious, impartial consideration by setting aside one's commitment towards one's own beliefs and perspectives. Learning to prepare and teach open-minded lessons is a crucial skill for student teachers because it fosters an atmosphere in which pupils feel free to express their own views and to learn about the views of others. The aim of this experiment was to examine which instructional strategy best supports student teachers' learning to prepare an open-minded citizenship education lesson. Therefore, participants (n = 176) processed an instruction on how to prepare an open-minded citizenship education lesson through learning by teaching on video (TOV), preparing to teach (PTT), or restudy (control condition), and as a posttest designed a lesson plan. We examined the completeness and accuracy of the explanations of the instructional content, feelings of social presence and arousal, open-mindedness levels, the completeness and accuracy of the lesson plans, and the conceptual knowledge of the instructional content. In addition, the lesson plans were graded on overall quality. Results showed that all participants scored higher on open-mindedness as measured with the Actively Open-minded Thinking Scale after the experiment than before the experiment. Participants in the control condition prepared a significantly more accurate and complete open-minded lesson then participants in the other two conditions, suggesting they have gained better understanding of the instructional content. There were no significant differences between the conditions on the other outcome measures.

Keywords: Teaching on video, generative learning, social presence, student teachers, open-mindedness, confirmation bias

Introduction

Learners often rely on superficial and inefficient learning strategies that target rote learning, and do not lead to meaningful knowledge in the long term (Dirkx et al., 2019; Hartwig & Dunlosky, 2012; Karpicke et al., 2009; Dunlosky et al., 2013; Fiorella & Mayer, 2016). It is therefore considered important to encourage learners to process learning material through generative learning strategies. Generative learning strategies contribute to effective learning because learners are actively involved in making the to-be-learned information meaningful by reorganizing and integrating this information with prior knowledge, which facilitates the transfer of what they have learned to new contexts (Fiorella & Mayer, 2016). The transfer of learning is the ultimate educational goal, i.e., learning beyond the initial course, task, or test (Barnett & Ceci, 2002; Lobato, 2006).

Generative learning strategies such as preparing to teach and learning by teaching on video are effective for knowledge building and have more robust effects on memory and transfer of knowledge compared to more superficial learning strategies, such as (massed) re-study (e.g., Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Hoogerheide, Renkl, et al., 2019; Kobayashi, 2019; Renkl, 1997; Roscoe & Chi, 2008). Preparing to teach means that learners study the learning material with a teaching expectancy by actually preparing a lesson on paper (Kobayashi, 2019; Muis et al., 2016). This implies that learners study the material and prepare a lesson while keeping in mind that he/she has to explain it at a later moment to some-one else. By doing so, the construction of deeper meaning of the concepts is enhanced compared to the often used learning strategies such as re-reading or highlighting (Dunlosky et al., 2013; Fiorella & Mayer, 2016). This is because the "teacher" benefits from explaining to others because one has to select the relevant information to include in the explanation, organizing it in a way that it can be understood by others, and one has to elaborate on the material by incorporating one's existing knowledge, which leads to new knowledge structures (Duran, 2017; Fiorella & Mayer, 2016).

The learning gains of preparing to teach might be strengthened by the actual act of teaching on video to a (fictitious) audience with the goal of helping others to learn (Fiorella & Mayer, 2013, 2016; Hoogerheide et al., 2016). Teaching on video presumably evokes feelings of social presence. Social presence can be defined as the awareness of a (fictitious) audience (Hoogerheide, Renkl, et al., 2019). Feelings of social presence might in turn generate arousal which may result in subsequent better learning and transfer compared to re-study (Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019). Indeed, the "learning by teaching on video" -

strategy has shown promising effects on learning and transfer across various ages, different learning materials and in various domains compared to restudying the learning material (Fiorella & Mayer, 2013; Hoogerheide et al., 2016; Hoogerheide et al., 2014a; Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019).

Learning by teaching on video may support student teachers' learning to prepare open-minded citizenship lessons. When a primary education teacher prepares a citizenship education lesson that addresses a topic that might provoke discussion in the classroom, e.g., racism, it is important to be open-minded. Open-mindedness is a crucial critical thinking disposition, and it is defined as one's willingness and ability to consider opposing experiences, beliefs, values, and perspectives and give them a serious, impartial consideration by setting aside one's commitment towards one's own experiences, beliefs, values and perspectives (Baehr, 2011; Facione, 1990; Kwong, 2016). By being open-minded during citizenship education, a teacher provides students with a good example of a consideration mode. In addition, it creates an atmosphere in which pupils feel free to express their own views and, hence, to learn about the views of others. Therefore, when preparing a citizenship education lesson, student teachers must hold the goal of an "open-minded lesson" closely in mind. That is, they have to prepare a lesson, which will allow students to express different perspectives to a social topic such as racism or sexual orientation.

Learning by teaching on video has been shown a beneficial strategy e.g., to learn to reduce the confirmation bias (Van Brussel et al., 2021), and acquiring problem-solving skills from worked examples (Hoogerheide, Renkl, et al., 2019). It is an open question whether learning by teaching on video supports student teachers' learning in the context of preparing open-minded citizenship lessons. To prepare an open-minded lesson, it is necessary that student teachers have knowledge about the concept of open-mindedness, the confirmation bias, and the designing principles of open-minded lessons. A fallacy that might hinder designing an open-minded lesson is the confirmation bias (Cavojova et al., 2018; Nickerson, 1998; Schwind et al., 2012; Stanovich et al., 2016; Sternberg & Halpern, 2020). The confirmation bias refers to the finding that people tend to be selective in finding and using evidence that is consistent with their own beliefs or expectations rather than selecting and processing inconsistent information (Cavojova et al., 2018; Nickerson, 1998; Schwind et al., 2012; Stanovich et al., 2016; Sternberg & Halpern, 2020; Tversky & Kahneman, 1974). As a consequence, the confirmation bias may result in one-sidedness where perspective taking is an important element, such as in citizenship education (e.g., Abrami et al., 2015; Nickerson, 1998; Schwind et al., 2012).

When a primary education teacher prepares a citizenship education lesson that addresses a topic that might provoke discussion in the classroom, e.g., racism, it is important to be open-minded. Open-mindedness is a crucial critical thinking disposition and it is defined as one's willingness and ability to consider opposing experiences, beliefs, values, and perspectives and give them a serious, impartial consideration by setting aside one's commitment towards one's own experiences, beliefs, values and perspectives (Baehr, 2011; Facione, 1990; Kwong, 2016). By being open-minded during citizenship education, a teacher provides students with a good example of a consideration mode. In addition, it creates an atmosphere in which pupils feel free to express their own views and, hence, to learn about the views of others. Therefore, when preparing a citizenship education lesson, student teachers must hold the goal of an "open-minded lesson" closely in mind. That is, they have to prepare a lesson, which will allow students to express different perspectives to a social topic such as racism or sexual orientation.

Learning by teaching on video is a promising strategy to gain meaningful knowledge about the confirmation bias and perspective taking, which student teachers have to apply to prepare open-minded citizenship lessons. In addition, this strategy may have a stronger effect on student teachers' open-mindedness compared to learning by preparing to teach and re-studying.

The present study[9]. The central question of the present study was which instructional strategy supports student teachers best to prepare an open-minded citizenship education lesson. To address this question, teaching on video was pitted against preparing to teach, and restudy. Participants were student teachers who first completed the Actively Open-minded Thinking scale (AOT; Stanovich & West, 2007). After one week all participants received an instruction[10] on substantive and procedural knowledge regarding open-minded lessons, confirmation bias and perspective taking. Subsequently, participants were assigned to one of the three conditions. Participants in the first condition, Teaching on video (TOV), processed the instructional content through preparing an explanation about the instructional content and taught this explanation in a video to a fictitious audience. Participants in the second condition, Preparing to teach (PTT), processed the instructional content through preparing an explanation of that content. Participants in the third condition, the control condition (CC), processed the instructional content by

9 The present study was pre-registered at the Open Science Framework: Hypotheses, planned data-collection, planned methods, planned analyses and the datasets generated during and analysed during the current study are available on the Open Science Framework

10 The instructional text in Dutch can be found on the Open Science Framework

restudying the text for 10 minutes. During the learning phase, feelings of social presence and arousal were measured through questionnaires in the teaching on video and preparing to teach condition[11]. After the learning phase, all participants completed the AOT for the second time. As a posttest, all participants wrote a lesson plan for a topic within the context of citizenship education lesson in primary school (e.g., racism, obesity). Finally, all participants completed a conceptual knowledge test to assesses their knowledge about important concepts from the instruction, e.g., confirmation bias and open-mindedness.

Hypotheses. In the present study, the quality of the explanation of the instructional content and the open-mindedness of the lesson plan were the most important variables. We hypothesized that the quality of the explanations in the learning phase and the degree of open-mindedness of the lesson plan after the learning phase, would be higher in the group who practised through teaching on video (TOV) compared to participants who only prepared to teach (PTT). This hypothesis is based on the follow line of reasoning: Participants who learn through teaching on video often experience more feelings of social presence which induce higher arousal levels because of addressing an audience (i.e., the social presence hypothesis; Gunawardena, 1995; Hoogerheide, Renkl, et al., 2019). As a result, participants are stimulated to generate accurate explanations to ensure that the audience understands the subject. Based on the generative learning hypothesis (e.g., Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019), we thus expected that participants in the TOV condition obtained a deeper conceptual understanding of what open-mindedness is and why it is important compared to respectively participants in the PTT and control condition (CC). Compared to TOV and PTT, participants in the control condition only re-study the instructional content. Compared to the generative learning strategies, re-studying does not involve deep processing of to-be learned content (Dunlosky et al., 2013; Fiorella & Mayer, 2016). If we assume that deep conceptual understanding is needed to prepare an open-minded lesson, then it is reasonable to assume that the accuracy and completeness of the open-minded lesson plan was highest in the TOV condition, followed by the PTT condition, in which accuracy and completeness will be higher than in the control condition, i.e., TOV > PTT > CC.

For conceptual knowledge, based on the idea of generative learning as outlined earlier in the Introduction, we hypothesized that participants who learned through teaching on video would perform better on the conceptual knowledge test compared

11 Due to a technical error, contrary to the pre-registration, the arousal questionnaire was not administered in the control condition.

to participants who only prepared to teach, and subsequently to participants who re-studied: TOV > PTT > CC for conceptual knowledge.

Furthermore, we exploratively compared mean AOT pretest to posttest scores within and between the conditions. Because the attitudes of the participants will not change quickly during a relatively short intervention, we will have to interpret the results cautiously.

Finally, learning by teaching on video probably induces higher feelings of social presence and arousal. Therefore, we exploratively compared the TOV and PTT conditions on self-reported feelings of social presence and arousal.

Method

Participants and design. To determine the required sample size for a standard sensitivity of the test procedure (i.e., power) of .80 for the One-Way (single factor) ANOVA, under a significance level of .05, and a medium effect size (f = .25), we needed to test at least 159 participants according to G*Power (Faul et al., 2007). This criterion was met, because 176 Dutch student teachers (M_{age} = 21.60, SD = 4.99, 153 women) from six Dutch primary school teacher education institutions participated in our research. At the time of the experiment, the concepts of confirmation bias and open-mindedness were not yet taught in the curriculum. Participants gave informed consent prior to the experiment. The first 160 participants were randomly assigned to the three conditions of the experiment. While the experiment was in progress, 16 additional participants registered for participation in the study. These took part in the control condition. Therefore, the distribution of the participants was as follows: TOV (n= 51), PTT (n = 54), and CC (n = 71). Participants either received course credits or a shop voucher. The rewards were not correlated with condition. The dependent variables were the quality of the explanation (TOV and PTT) and the open-mindedness of the lesson plan, i.e., the degree to which the lesson plan contained multiple perspectives on the topic at hand, and whether it left room for an open-minded discussion. Feelings of social presence and arousal (TOV and PTT conditions only), conceptual knowledge, and the tendency towards open-minded thinking were also measured.

We pre-registered the hypotheses and shared the data on the Open Science Framework. By preregistering and storing all data on the Open Science Framework, we refrained from biasing the results by e.g., null hypothesis significance testing (i.e., NHST), or p-hacking. Our view is that preregistration and open science are important ways to achieve more transparency and objectivity in science (e.g., Conlin et al., 2019; Munafò et al., 2017; Nosek & Lakens, 2014; Simmons et al., 2011).

Materials

All materials were delivered in Dutch through the online Qualtrics platform (Qualtrics, 2017). In the Teaching on video condition, the actual act of teaching was recorded on participants' smart phones and sent to the researcher via WhatsApp or e-mail. The participants could not click back to previous parts during the experiment.

Instruction. The text-based instruction consisted of approximately 1800 words. The instructional content concerned the concepts of critical thinking, confirmation bias, open-mindedness and perspective taking and the steps that are needed to design an open-minded lesson were described. Some examples and didactical and pedagogical suggestions (hints) were provided to further explain the importance of having an open-mind and how to prepare a citizenship education lesson on a social topic.

Conceptual knowledge test. The conceptual knowledge test consisted of six open questions about the content of the instruction, e.g., "Explain what the confirmation bias is". This test was designed by the first author and was aimed at testing participants' knowledge about the concepts that were addressed in the instruction (See Appendix A).

Open-mindedness. The Dutch version of the Actively Open-minded Thinking scale (translated in Dutch by Heijltjes et al., 2014; Stanovich & West, 2007) was used to measure participants' open-mindedness. This scale is aimed at measuring the level of one's open-minded thinking as a thinking disposition. The test consists of 41 items to which participants have to respond on a 6-point Likert scale, ranging from (1) strongly disagree to (6) strongly agree. Higher scores on the AOT imply a greater tendency towards open-minded thinking. Lower scores indicate closed-minded thinking which leads to e.g., the confirmation bias in reasoning and decision making (Baron, 2008; Stanovich & West, 2007). In general, studies that use the AOT, report a high reliability of the test (For an overview, see Janssen et al., 2020, pp. 2, Table 1).

Examples of AOT items are: "I believe that the different ideas of right and wrong that people in other societies have may be valid for them", "Someone who attacks my beliefs is not insulting me personally". Some items have to be reversed before analysis, e.g., "I tend to classify people as either for me or against me".

Social presence. The social presence questionnaire with 10 statements was constructed by Van Brussel et al. (2021) and in their study the Cronbach's alpha was .73. In the social presence questionnaire (See Appendix B), participants had to indicate on a scale from (1) strongly disagree to (5) strongly agree to which degree each of the ten statements represents how they felt after the learning phase. We used sum scores per participant: Scores run from 10 (10 x 1), to 50 (10 x 5). The higher the score, the higher the feelings of social presence.

Arousal. To measure arousal, the activitation-deactivation adjective check list (ADACL, see Appendix C) by Thayer (1967, 1986) was used. This checklist was translated to Dutch and used in a former study by the authors (Van Brussel et al., 2021). It assesses core arousal or activation states based on two dimensions (activation and deactivation) and four subscales i.e., energetic, tiredness, tension and calmness. Each subscale consists of five adjectives. Participants rate on a four-point scale how well the adjective described their immediate feelings after the explanation phase (4 = "definitely feel", 3 = "feel slightly",2 = "cannot decide", and 1 = "definitely do not feel this way"). Per subscale we averaged the scores of the five adjectives. "Wakeful" and "wide-awake" were reversed for the Tiredness subscale. Higher scores indicated higher levels of arousal. Previous studies revealed excellent Cronbach's alphas on all four categories (Boyle et al., 2015; Thayer, 1978). Van Brussel et al. (2021) found a Cronbach's alpha of .82 on "energy", and .82 on "tiredness", but questionable Cronbach's alphas on both "tension" (.67) and "calmness" (.60).

Procedure. Participants were tested online because due to the restrictions of the Covid-19 crisis in 2020, they were not allowed to attend the university building. They were called upon to work individually, focused and without disturbance. See Figure 1 for a visualization of the procedure.

One week before the experiment, participants completed the AOT for the first time. After one week, participants received an e-mail with the link that led to the Qualtrics platform. Participants read the instructions and by clicking to continue, they gave informed consent to use their data for the research. The experimenter was available by phone or e-mail throughout the experiment for practical questions. All participants started with studying the instruction. Next, the three groups each followed their own intervention to process the instruction.

In the teaching on video group (TOV), participants received the following instruction: "Prepare an explanation of what you have just learned and then provide this explanation to your peer student teachers who are not participating in this project. You do this by recording the explanation via the camera of your smartphone. They will watch your explanation online at a later time. It is therefore important that you give an accurate and complete explanation". Participants had to write down their preparation in Qualtrics before the teaching started, and they were told that their video would be used for online activities which are currently common in the Netherlands during the Covid-19 crisis to create the most authentic situation as possible[12]. The recording of the explanation was sent by e-mail or WhatsApp to the experimenter. In the preparing to teach group (PTT), wrote a preparation of their explanation for their peers in Qualtrics. Participants in the TOV and PTT conditions were called upon to set the time for this phase for 10 minutes. In the re-study control condition (CC), participants re-studied the instruction for 10 minutes after which they were automatically forwarded to the next page in Qualtrics.

After this, TOV and PTT conditions completed the social presence and arousal questionnaire and all three conditions (TOV, PTT, and CC) completed the Actively Open-minded Thinking scale for the second time. Then, all participants, including the control condition, received the assignment to prepare a lesson plan for a citizenship education lesson for primary school children in 6th grade (11–12-year-old children) on a topic that can provoke discussion: "You will be teaching on one of the themes below. Choose a theme and write your lesson plan, be complete and accurate so that others can also teach your lesson. Be concrete: What do you say, what do you do, what do you ask? The themes are: Radicalization of young people, migration and refugees, LGBTI-youth, religion / belief, childhood obesity, mouth masks in public transport[13] and (black) Pete. Please note, there is a minimum number of characters that you must use. If you can't click to the following page yet, you'll have to explain more.". In their lesson plan, participants were supposed to show open-mindedness by taking perspective on the chosen social topic through considering opposites and alternatives to their own point of view without giving the explicit

12 During the debriefing, TOV participants were told that this was a cover-up story and that their recordings were safely stored and would be destroyed after the analyses.

13 This research was conducted at the beginning of the Covid-19 pandemic in 2020 and in the Netherlands, it was not yet mandatory to wear mouth masks everywhere, except in public transport. Therefore, at the time of the experiment, this was a debatable topic.

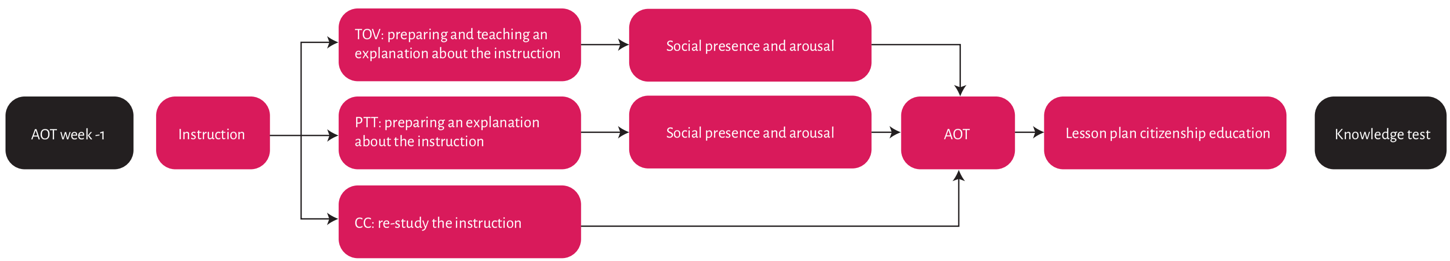

Figure 1: Procedure of the experiment

instruction to do so. The minimum amount of characters was 1250. Participants were not able to continue until they reached that amount of characters. This was to prevent participants from rushing the assignment. None of the conditions had the instructional text at hand.

All participants then made the conceptual knowledge test. After the test, they were asked for their prior education, sex and age. Finally, participants were thanked and given the opportunity to receive a summary of the research results.

Data analysis

Based on our data analysis plan of the pre-registration, we checked variables on missing data and outliers. There we no missing data. Prior to running a statistical test, we checked for multivariate and univariate outliers. We ran and reported all analyses with and without outliers, but outliers – if any - did not change our results. For the AOT, conceptual knowledge test and the social presence and arousal questionnaires, a minimal Cronbach's alpha level of .70 was set as a threshold for the analysis of a sum score or average score.

Qualtrics automatically reported the submission time of the instruction page. We observed five outliers at the upper side of the reading time for the instruction. Without these outliers, the mean time spent on reading the instruction was 9.45 minutes (SD = 11.78 and Median = 8.20 minutes). Analyses without these participants did not influence the results, therefore we decided to include the data of all participants.

Analysis of the explanation and lesson plan. The explanation (TOV, n = 51 and PTT, n = 54), and the open-minded lesson plan (TOV, PTT, and CC, n = 71) were analysed. An independent rater and the first author scored 30% of the explanations and 28% of the lesson plans to check for the reliability of the scoring method. The interrater reliability was "substantial" according to the interpretation of Cohen's kappa (Cohen, 1960; Landis & Koch, 1977): completeness explanation: $\kappa = .73$, accuracy explanation: $\kappa = .73$, completeness lesson plan: $\kappa = .71$, and accuracy lesson plan: $\kappa = .79$. Therefore, one rater scored the remainder of the texts and these results were used in the analyses.

The quality of the explanation and the open-mindedness of the lesson plan were operationalized by scoring items on completeness and accuracy. See Appendix D for the scoring forms. The explanation (TOV and PTT) was firstly assessed on the presence of six concepts that were addressed in the instruction (e.g., explaining confirmation bias). This resulted in a completeness score. Next, the explanation of the concepts present, were scored on the accuracy of the explanation (e.g., the explanation of the confirmation bias was accurate).

The lesson plan (TOV, PTT, and CC) was firstly assessed on whether the instructional content was incorporated in the lesson. One of the four items was for example "The explanation shows that the teacher presents multiple perspectives with regard to how people can think about the theme (= applying perspective taking to avoid the confirmation bias / showing open-mindedness)". This led to a completeness score. Subsequently, the accuracy of the present incorporated content was scored.

Completeness was measured through the presence (1 point) or absence (0 points) of concepts (e.g., a participant earns 1 point if the explanation of open-mindedness is present). For each accuracy item, the answer rate was correct (1 point) partly correct (0.5 points) or incorrect (0 points). (e.g., there were missing elements in the explanation of open-mindedness, so the participant earned 0.5 points). The approach for this analysis was based on Hoogerheide, Renkl, et al. (2019) and Van Brussel et al. (2021). The maximum score for the explanation was 10 points per category, and for the lesson plan 4 points per category. For the analyses, we conducted One Way ANOVAs.

Beyond our pre-registration, we exploratively conducted an overall evaluation of the quality of the lesson plan because retrospectively, in our view, the pre-registered scoring was quite narrowly focused on the exact instructional content. Therefore, it probably left elements underexposed that indicated that a participant learned to prepare an open-minded lesson. The first author, who is an experienced teacher educator and assessor, scored the lesson plans blinded for condition on the following criteria: 1) In general, the lesson is aimed at stimulating open-mindedness towards the topic (i.e., open-mindedness as defined in the Introduction), 2) The content of the lesson plan shows that elements of the instruction are applied in the lesson plan, and 3) Teaching or working methods that contribute to open-mindedness are described. The lesson plans ($n = 176$), were scored based on the Dutch rating system in which assessment scores range between 1 (very insufficient) and 10 (excellent). We conducted a non-parametric Kruskal-Wallis Test on the overall quality scores.

Analysis of the conceptual knowledge test. For each of the six items of the conceptual knowledge test, a maximum score of six points could be obtained.

Participants earned 1 point (accurate), 0,5 point (partly accurate) or no points (wrong answer). The test was, however, not reliable: Cronbach's alpha = .41. Therefore, contrary to what was preregistered, we conducted explorative One-Way ANOVAs on the individual items.

Analysis of open-mindedness. Negatively formulated AOT items were reversed as indicated on the test form by Heijltjes et al. (2014). The AOT was individually scored on both test moments: for each participant we calculated the mean scores of the 41 items. The initial measurement had a low Cronbach's alpha of .53. The test was reliable on the second measure: Cronbach's alpha = .83. To determine whether there were differences between the three groups, we conducted a 3 (Condition) x 2 (Pretest, Posttest) mixed ANOVA.

Analysis of the social presence questionnaire. The analysis for this questionnaire was based on the approach by Van Brussel et al. (2021). Items 4, 9 and 10 were reversed before analysis. To achieve the minimum required Cronbach's alpha level of .70, items 3 and 10 were deleted, which resulted in a Cronbach's alpha of .70. Analyses were conducted with the subset of 8 items. To explore any differences on feelings of social presence between the two conditions, an independent samples t-test was conducted on the sum scores per participant.

Analysis of the arousal questionnaire. The ADACL was used to investigate whether arousal level differences between the conditions could explain different effects of the two instructional strategies. "Wakeful" and "wide-awake" were reversed for the Tiredness subscale. For the Tension subscale, after deleting item 6, the Cronbach's alpha was .83. After deleting the first item, the Energetic subscale showed a Cronbach's alpha of .76. However, the other two subscales were not reliable: For Tiredness, even after deleting item 3, Cronbach's alpha was .65, and for Calmness the Cronbach's alpha was only .61 after deleting item 14. Therefore, explorative analyses of these subscales were conducted on individual items Mean scores per subscale per participant were calculated. For the Tiredness and Tension subscales and the items of the subscales Energetic and Calmness, we conducted independent samples t-tests to determine differences between conditions.

Results

In the control condition, 27% of the participants indicated that they re-read the text once, 61% re-read the text two times and 12% reported "other" (e.g., 3 times, or 1.5

times). In all analyses below, a significance level of .05 was used as a threshold for statistical significance. Eta-squared (η^2) is reported as measure of effect size for the ANOVAs for which .01 is considered small, .06 medium, and .14 large.

The explanation and lesson plan. The 105 participants in the TOV and PTT conditions prepared an explanation after the intervention as part of the instructional strategy. All 176 participants prepared an open-minded lesson plan as a posttest. See Table 1 for the relevant descriptive statistics.

Table 1. Mean Scores and Standard Deviations, per Condition for the Explanation and the Lesson Plan

	Explanation (max. 10 points)		Lesson Plan (max. 4 points)	
	Completeness	**Accuracy**	**Completeness**	**Accuracy**
	M (SD)	**M (SD)**	**M (SD)**	**M (SD)**
TOV (*n* = 51)	4.24 (2.04)	3.56 (1.97)	1.59 (1.24)	1.17 (1.01)
PTT (*n* = 54)	3.89 (2.51)	3.48 (2.05)	1.69 (1.32)	1.26 (1.20)
CC (*n* = 76)	N/A	N/A	2.16 (1.38)	1.80 (1.02)

The explanation (TOV and PTT). The conditions neither differed significantly on completeness: F(1,103) = .71, p = .40 with a small effect size of η^2 = .007, nor on accuracy: F(1,103) = .12, p = .73, again with a small effect size of η^2 = .001. Two outliers for accuracy were detected but running the analyses without them did not yield other results: F(1,101) = .049, p = .825, η^2 = .001.

The lesson plan (TOV, PTT and CC). All participants wrote a lesson plan after the intervention and there were no outliers. Participants chose one of the given topics for the lesson plan: Radicalization of young people (1.1%), Migration and refugees (5.7%), Religion (6.3%), Childhood obesity (12.5%), LGBTI community (21%), Mouth masks in public transport (22.2%), and Black Pete (31.3%). Participants who chose Black Pete scored highest on completeness and accuracy, whereas participants who chose Radicalization of young people scored lowest on both variables.

There were significant differences between the conditions on the completeness of the lesson plan, F(2,173) = 3.32, p = .039, η^2 = .037, and for accuracy, F(2,173) = 5.05, p = .007, η^2 = .055. As a follow up, planned Helmert contrast were performed. The contrast showed that the completeness of the lesson plan in the TOV condition was not significantly lower with a contrast estimate of -.33 (SE = .22), p = .133, compared to the combined completeness score of PTT and CC. A contrast estimate of -.47

(SE = .24), $p = .050$ showed that PTT participants scored lower on completeness than CC participants. For accuracy, the contrast showed that the lesson plan in the TOV condition was not significantly lower with a contrast estimate of -.36 (SE = .20), $p = .073$ compared to the combined accuracy score of PTT and CC. A contrast estimate of -.54 (SE = .22), $p = .014$ did reveal that PTT participants scored lower on accuracy compared to CC participants. In sum, participants who re-read the instruction delivered a more complete and accurate lesson plan compared to participants who processed the instruction through explaining and to participants who taught on video. In sum, participants who re-read the instruction delivered a more complete and accurate lesson plan compared to participants who processed the instruction through explaining and to participants who taught on video.

For the analysis of the overall quality, the mean score for the TOV lesson plans was 7.10 (SD = 2.20), for the PTT lesson plans 7.00 (SD = 2.05), and for the CC lesson plans the mean score was 6.87 (SD = 2.08). The results showed however, that the three groups did not differ significantly from each other on the quality of the lesson plan ($H(2) = .311$, $p = .856$). See Table 2 for percentages per condition per ordinal scale: insufficient, sufficient, good, and excellent.

Conceptual knowledge test. Since the test was not reliable ($\alpha = .41$), results are presented on item level in Table 3. We conducted explorative One-Way ANOVAs which revealed significant differences on item 3: $F(2,173) = 4.723$, $p = .010$, and item 5: $F(2,173) = 13.776$, $p < .001$. To find out which conditions differed, independent t-tests were conducted on these two items (see Table 3).

Table 2. Scores of Overall Quality of the Lesson Plan per Condition as Percentages per Scale

	Condition		
	TOV	***PTT***	***CC***
Scale	***n* = 51**	***n* = 54**	***n* = 71**
Insufficient	20%	19%	21%
Sufficient	33%	33%	34%
Good	20%	28%	25%
Excellent	27%	20%	20%

Table 3. Conceptual Knowledge Test Scores and Standard Deviations per Proportion Correct per Item per Condition

	TOV *n = 51* *M (SD)*	PTT *n = 54* *M (SD)*	CC *n = 76* *M (SD)*
1. The definition of critical thinking	0.28 (.39)	0.19 (0.39)	0.28 (0.38)
2. Explanation of the confirmation bias	0.74 (0.34)	0.78 (0.33)	0.74 (0.35)
3. A characteristic of an open-minded person	0.87 (0.28)[A]	0.80 (0.33)[A/B]	0.68 (0.39)[B]
4. Steps that contribute to preparing a confirmation bias-free explanation	0.76 (0.31)	0.70 (0.36)	0.70 (0.36)
5. Explain the importance to reduce the confirmation bias	0.76 (0.32)[A]	0.82 (0.33)[A]	0.51 (0.39)[B]
6. Case and an adequate respons	0.47 (0.44)	0.33 (0.40)	0.53 (0.68)

Note. Significant differences are indicated with A and B (independent t-test, $p < .001$)

Open mindedness. We conducted a 3 (Condition: TOV, PTT, CC) x 2 (Pretest vs. Posttest) Mixed ANOVA with Condition as a between-subjects factor on the AOT scores. In Table 4, the mean individual scores and SDs on both test moments are presented. We found a main effect of Test Moment: $F(1,173) = 3515.33$, $p < .001$, $\eta^2 = .95$: all participants scored higher on the second measurement compared to the first. However, we found no main effect of Condition: $F(2,173) = .049$, $p = .952$. Also, no interaction effect was found; $F(2, 173) = .19$, $p = .824$ and $\eta^2 = .002$.

Table 4. Mean Individual Item Scores and Standard Deviations of the Actively Open-minded Thinking Tests per Condition

	AOT1 *M (SD)*	AOT2 *M (SD)*
TOV n = 51	2.93 (0.17)	4.42 (0.37)
PTT n = 54	2.92 (0.20)	4.44 (0.37)
CC n = 76	2.93 (0.18)	4.45 (0.35)
Total mean	2.93 (0.18)	4.44 (0.36)

Social Presence. The mean score for the TOV participants on the 10-item social presence questionnaire was 29.24 (SD = 4.51), and for the PTT participants 30.56 (SD = 4.51). This difference was not significant: $t(103) = -1.50$, $p = .137$.

Arousal. See Table 5 for the descriptive statistics of the ADACL subscales for the TOV and PTT condition.

An independent samples t-test showed that for the subscale Tension, a significant difference with higher mean scores for the PTT participants was found: $t(103) = -2.541$, $p = .013$. No significant difference was found for Energetic $t(103) = -1.461$, $p = .147$. Conducting the analyses without outliers, did not yield other results. For the subscales Tiredness and Calmness, the analyses on individual items only revealed a significant difference between TOV and PTT on item 14 "still" of the Calmness subscale; $t(103) = 3.636$, $p < .001$ with a higher mean score for TOV ($M = 3.35$, $SD = .98$) compared to PTT ($M = 2.61$, $SD = 1.11$).

Table 5. Mean Scores and Standard Deviations of the Subscales of the ADACL

	Tension	**Energetic**	**Tiredness**	**Calmness**
	M (SD)	***M (SD)***	***M (SD)***	***M (SD)***
TOV (n = 51)	2.96 (0.81)	2.02 (0.69)	3.45 (0.60)	2.30 (0.67)
PTT (n = 54)	3.35 (0.74)	2.21 (0.63)	3.33 (0.58)	2.13 (0.60)

Conclusion and Discussion

The aim of this experiment was to examine whether the instructional strategy affects the open-mindedness of student teachers' lesson plans on social topics within citizenship education at primary school. An instruction on confirmation bias, perspective taking and open-mindedness, and the role of these concepts when preparing an open-minded lesson constituted the basis for three instructional strategies: Teaching on video, Preparing to teach, and Re-study. In contrast to our hypothesis, the results showed that teaching the instructional content to a fictitious peer audience did not lead to more complete and accurate explanations compared to only preparing. As a posttest, participants in the control condition designed an open-minded lesson plan that was more accurate and complete compared to the other two conditions. An analysis of the overall quality of the lesson plan revealed no significant differences between the three conditions. All participants showed progress on the Actively Open-minded Thinking Scale after the intervention. We will discuss these findings and the limitations of the study.

We expected that participants in the TOV condition would score higher on social presence compared to PTT participants. There were, however, no significant differences between the conditions on feelings of social presence. A potential explanation might be the timing of the measurement. We measured feelings of social presence after the actual act of preparing and teaching which may not have represented the feelings of social presence during the task and is a limitation of the current study. Therefore, feelings of social presence might have faded away because the actual act of teaching was over. Other measures such as wristbands that measure a direct change in the electrical resistance or temperature of the skin caused by e.g., arousal (e.g., Biocca & Harms, 2002; Cui, 2013; Gunawardena & Zittle, 1997; Hoogerheide, Renkl, et al., 2019), or counting the number of personal references in the explanation (e.g., Hoogerheide et al., 2016; Jacob et al., 2020; Lachner et al., 2018) might have yielded other results. For follow-up research, it is interesting to use both types of measurements and measure during and after the intervention.

Another explanation as to why the TOV and PTT conditions did not differ, might be that participants did not receive enough cues indicating that their video would actually be watched by a peer. Therefore, they might not have really believed that peers would watch their explanation. This might have attenuated their feelings of social presence and arousal and hence, this might have worked against an additional effect of teaching on video relative to preparing to teach. However, in the study by Hoogerheide, Renkl, et al. (2019), which used a similar instruction in the TOV condition as we did, and in which a real audience was also absent, participants in the TOV condition did show higher levels of arousal than participants in the control condition who only studied worked examples. Hence, it is not evident that the cues that we used in our instruction were ineffective in leading participants to believe that their videos would be used for a real audience.

As a posttest, all participants prepared an open-minded lesson on a social topic. The results were, however, not consistent with our predictions, because participants who re-read the instruction created a more complete and accurate lesson plan compared to the participants in the TOV and PTT condition. An explanation of this result might be that the TOV and PTT participants experienced more mental effort because they had to retrieve the information of the instruction from their working memory during the learning phase in which they prepared, or prepared and taught an explanation about the instructional material (Paas, Tuovinen, et al., 2003; Van Gog et al., 2015). Participants in the control condition, did have the learning material at hand to re-study. In the teaching on video study by Hoogerheide, Renkl, et al. (2019), participants had to teach a worked example to peers with the example at hand. They reported higher effort investment compared to participants in a control condition who had to study the example. Teaching the example also led to better scores on a posttest with problem solving tasks, which is an indication that their perceived effort investment was beneficial for learning. In our study however, TOV and PTT participants did not have the learning material at hand. They had to retrieve the learning material from their working memory. Therefore, the re-study control condition, who had the instructional content at hand, when they processed the instruction, compared to the re-study control condition. TOV and PTT participants had to retrieve the learning material from their working memory during the learning phase. Therefore, the re-study control condition, who had the instructional content at hand during the learning phase, probably scored better, at least on the immediate posttest that we administered.

Another explanation of the absence of performance differences might have been the quality of the explanations of the TOV and PTT participants. This quality was not particularly high in both conditions in both experiments. Effects of self-explanation on learning and performance are contingent on a sufficiently high quality of self-explanations. It might be possible, that the level of understanding participants reached after instruction and practice tasks was not high enough to allow for beneficial effect of teaching on video and preparing to teach to emerge (e.g., Jacob et al., 2021). All in all, it is still not fully clear when increased social presence is desirable or not for learning (Oh et al., 2018), especially in the context of generative learning for novices (Jacob et al., 2021). Further research is needed to examine when and why feelings of social presence are beneficial for learning, and in which learning contexts.

However, in our study, TOV and PTT were in fact combined with retrieval practice and retrieval practice typically reveals its positive effect after a longer term on performance (Rawson et al., 2013; Roediger & Karpicke, 2006, but see Van Gog et al. (2015) for contrasting findings). In the short term, immediately after processing the learning material, often no differences are found between retrieval practice and a more superficial learning strategy. Therefore, it might be possible that we would have found a positive effect of TOV and PTT relative to the control condition, if we had used a delayed test, for example a test administered one week after the instruction. This suggest that the effect of TOV or PTT might depend on whether the generative strategy of preparing to teach or teaching on video is combined with retrieval practice. Future research might investigate whether the effect of TOV or PTT interacts with retrieval practice.

Another interesting finding of our study is that all participants became more aware of the importance of actively searching for opposing evidence against one's own beliefs and the ability to weigh the available evidence fairly as measured by the AOT. Higher scores on the AOT are positively related to considering more alternative possibilities than one's initial point of view (Baron, 2008; Stanovich &

West, 2007). Our finding is in line with earlier studies with comparable content in which considering multiple perspectives reduced participants' confirmation bias (Adame, 2016; Lord et al., 1984; Mussweiler et al., 2000; Van Brussel et al., 2020, 2021). An explanation for this finding might be that all participants received the same instruction on open-mindedness, and that the instruction and the demand characteristics of the subsequent tasks determined the response on the 41 AOT-items. There is, however, an ongoing debate whether the AOT measures open-minded thinking as a unidimensional trait notwithstanding its high Cronbach's alphas. Janssen et al. (2020), found that despite various scale and factor analyses, neither the 41-item AOT, nor a subset of items measured open-minded thinking as a unidimensional trait that could discriminate between participants. Therefore, we have to interpret the AOT scores cautiously. All in all, the present study showed that preparing an open-minded lesson plan is an ecological valid manner to measure student teachers' confirmation bias. These results might imply that extending this task with observing the actual teaching and the degree to which the confirmation bias in the classroom might enhance performance even more. This might be the focus of future research.

Finally, the significance of our study is that the results show that re-study has a positive effect on student teachers' learning i.e., designing an open-minded lesson plan compared to generative processing, All participants scored higher on the Actively Open-minded Thinking scale after the intervention. This is an important teacher disposition to refrain from one-sidedness in citizenship education lessons on debatable topics, or unbiased high-stake decision-making regarding pupils (Abrami et al., 2015; Vanlommel et al., 2018). It is, therefore, important to continue investigating learning by teaching on video to better support student teachers' learning to prepare and teach open-minded lessons.

APPENDIX A

Conceptual Knowledge Test

(Translated from Dutch)

1. What is the definition of critical thinking used by Avans University of Applied Sciences for its students and teachers?

2. Explain what the confirmation bias is.

3. What is a characteristic of an open-minded person?

4. Name at least two steps that contribute to preparing a confirmation bias-free explanation about a controversial topic to group 7/8.

5. Explain why it is important for children to learn to reduce their confirmation bias.

6. A student indicates in class that she is against wearing mouth masks in public transport because she believes it leads to anxiety. What is an adequate response to its position and its substantiation?

APPENDIX B

Social Presence Questionnaire

(Translated from Dutch)

Please indicate to which degree the following statements represent how you felt during the learning phase

(1) = strongly disagree
(2) = disagree
(3) = undecided
(4) = agree
(5) = strongly agree

1. I was aware of an audience / social presence during the explaining assignment [+]

2. I thought about what would happen if I did not understand the learning material [+]

3. I was motivated to accurately explain the learning material to my peers [+]

4. I did not feel engaged with my peer audience while explaining [-]

5. I was motivated to make a complete explanation for my peers [+]

6. I explained my lesson explicitly with my peer audience in mind [+]

7. My peer audience stimulated me to do the best I can [+]

8. I felt some tension while explaining because the learning material was rather complex [+]

9. That a peer audience will watch my explanation, does not bother me at all [-]

10. Teaching my peer audience, hampered my own learning [-]

APPENDIX C

Activation Deactivation Adjective Check List (ADACL, Thayer, 1986)

Each of the words here describe feelings or moods. Please use the rating scale next to each word to describe your feelings or mood at this moment. Work rapidly, but please mark all the words. Your first reaction is the best. This should take only a minute or two.

(1) = definitely feel | (2) = feel slightly | (3) = cannot decide | (4) = definitely do not feel

1. Active (subscale Energetic)
2. Placid (subscale Calmness)
3. Sleepy (subscale Tiredness)
4. Jittery (subscale Tension)
5. Energetic (subscale Energetic)
6. Intense (subscale Tension)
7. Calm (subscale Calmness)
8. Tired (subscale Tiredness)
9. Vigorous (subscale Energetic)
10. At-rest (subscale Calmness)
11. Drowsy (subscale Tiredness)
12. Fearful (subscale Tension)
13. Lively (subscale Energetic)
14. Still (subscale Calmness)
15. Wide awake (subscale Tiredness)
16. Clutched up (subscale Tension)
17. Quiet (subscale Calmness)
18. Full of pep (subscale Energetic)
19. Tense (subscale Tension)
20. Wakeful (subscale Tiredness

APPENDIX D

Scoring Forms for the Explanation (during practice) and Lesson Plan (posttest)

Explanation (TOV and PTT)	**Completeness** Present = 1 Missing = 0	**Accuracy** Accurate description = 1 Partly accurate description = 0.5 Wrong description = 0
1. The teacher explains the concept *critical thinking*		
2. The teacher explains the concept *confirmation bias*		
3. The teacher explains the concept *open-mindedness*		
4. The teacher explains the concept *perspective taking*		
5. The teacher explains the three steps to prepare a confirmation bias-free lesson	Three steps = 3 Two steps = 2 One step = 1 No step = 0	
6. The teacher has incorporated the five suggestions of the instruction into the explanation	Five suggestions present = 3 Four suggestions = 2.5 Three suggestions = 2 Two suggestions = 1.5 One suggestion = 1 No suggestion = 0	Per suggestion present (e.g., maximum score = 5 x 1 point)
Total score (maximum 10 per category)		

Lesson plan (TOV / PTT / CC)	**Completeness** Present = 1 Missing = 0	**Accuracy** Accurate description = 1 Partly accurate description = 0.5 Wrong description = 0
1. The explanation shows that the teacher presents multiple perspectives with regard to how people can think about the theme (= applying perspective taking to avoid the confirmation bias / showing open-mindedness)		
2. The teacher explains that it is important to also search for information that goes against someone's own point of view so that someone is better able to take a stand regarding the theme		
3. The teacher has formulated questions that show that she / he wants to make the pupils aware of the confirmation bias		
4. The teacher explains that it is also fine to change points of view if someone is convinced of a different perspective		
Total score (maximum 4 per category)		

Chapter 6

Summary and general discussion

The confirmation bias is a ubiquitous bias that regularly interferes with reasoning, decision making and judgment (e.g., Nickerson, 1998). In this thesis, we argued that mitigating the confirmation bias should be an important element of the primary teacher education curriculum, because it keeps teachers from making biased decisions and judgments regarding their pupils and teaching activities. An effective thinking strategy for bias reduction is to consider opposites and alternatives before making a decision or judgment (Adame, 2016; Hirt & Markman, 1995; Larrick, 2004; Lord et al., 1984; Mussweiler et al., 2000). The thinking strategy "consider-the-opposite" (COS) consists of asking oneself the question, "What are some reasons that my initial judgment might be wrong?" (Larrick, 2004, p. 323). It encourages people to direct attention to alternative evidence that would otherwise not be considered, thereby stimulating perspective taking.

The central goal of the research described in this dissertation was to investigate what instructional design principles are effective for student teachers to learn to mitigate their confirmation bias in hypothesis-testing, and to reduce their bias when designing a lesson plan for citizenship education lessons. Specifically, effects of learning context (Chapter 2), feedback (Chapter 3), and generative processing through learning by teaching on video (Chapter 4 and 5), were investigated. In all four studies, students participated in a pretest – intervention – posttest design. The intervention in each study consisted of an instruction and practice with confirmation bias tasks. In the first three studies, confirmation bias was operationalized as students' performance on confirmation bias tasks commonly used in the literature, i.e., hypothesis-testing tasks and four-card selection tasks (Jonas et al., 2001; Wason, 1968). In the fourth study, confirmation bias was operationalized as the level of open-mindedness displayed in a citizenship education lesson plan about a debatable social topic. Next to student teachers' performance on measures of confirmation bias, topic interest (Chapter 2), mental effort (Chapter 3), social presence and arousal (Chapter 4 and 5), and conceptual knowledge and open-mindedness (Chapter 5), were used as dependent variables.

In this Chapter, the summary of the findings is presented, and overarching issues and directions for future research, as well as methodological challenges are discussed. Finally, this chapter concludes with potential implications for educational practice.

Summary of the findings

In the first study, presented in **Chapter 2,** the central question was what the effect of instruction and practice with confirmation bias tasks in a domain-specific (educational) context is on student teachers' confirmation bias and their topic interest compared to instruction and practice with confirmation bias tasks in a domain-general context. After the confirmation bias-pretest, student teachers watched an instructional video about a "consider-the-opposite" thinking strategy for bias reduction. Subsequently, they practised confirmation bias tasks in either an educational context, or a general context and completed a topic interest questionnaire. Then they took the immediate confirmation bias-posttest. Delayed confirmation bias tests were administered one week and five weeks later. Results revealed that the mean confirmation bias score decreased from the pretest to the first delayed posttest, but that the mean decrease was similar in the two conditions. However, participants who learned and practised in the educational context reported higher topic interest compared to those in the general context condition.

In the study presented in **Chapter 3**, the central research question was whether elaborative (worked example) feedback would be more effective to learn to reduce the confirmation bias after instruction and practice tasks than correct answer feedback, or no feedback. Performance on confirmation bias tasks and transfer tasks and mental effort was measured. All participants completed a pretest in which confirmation bias was assessed with hypothesis-testing tasks and four-card selection tasks. Subsequently, five conditions were created. There were two feedback conditions, namely (1) a condition in which participants received a video-based "consider-the-opposite" instruction and practice tasks with elaborative feedback in the form of worked examples (WE, worked example), and (2) a condition in which participants received a video-based "consider-the-opposite" instruction and practice tasks with correct answer feedback (CA, correct answer). In the three other conditions, participants received no feedback, namely (3) a video-based "consider-the-opposite" instruction and practice tasks without feedback (PO, practice only), (4) a video-based "consider-the-opposite" instruction, without practice tasks (VO, video only) and (5) a no treatment control condition (NT, no treatment). Mental effort was measured during practice in the WE, CA and the PO condition. After this, participants completed the posttest consisting of confirmation bias learning items, i.e., items pertaining to tasks that were addressed during the instruction (VO condition) and practice tasks (WE, CA and PO condition) and transfer items. In the no treatment control condition (NT condition), participants took this posttest immediately after the pretest. The results showed that participants in both feedback conditions performed significantly better on isomorphic posttest tasks, i.e., the learning items, compared to the other three (no feedback) conditions, but no differences between the conditions were found on the transfer items. Except for the no treatment control condition, the mean score on the learning items in the other four conditions increased from pretest to posttest. Mental effort scores were rather low and did not differ as a function of condition.

The third experiment, presented in **Chapter 4**, examined the effect of learning by teaching on video on student teachers' confirmation bias. Therefore, teaching on video was pitted against preparing to teach, and a re-study control condition. The quality of the lesson preparation, confirmation bias, feelings of social presence, arousal, and transfer were measured. In this study, all participants took a confirmation bias-pretest, read the instruction and practised with tasks on which they received elaborative (i.e., worked example) feedback. After this, participants were asked to either prepare and teach a lesson on video about the instruction, only prepare a lesson about the instruction, or re-study the instruction. Afterwards, participants completed a social presence questionnaire, an arousal questionnaire, and a confirmation bias-posttest and one transfer-test task. Results showed that confirmation bias reduction was comparable in all conditions, but the quality of the lesson preparation was higher for the teaching on video participants than for the participants who only prepared a lesson (without a video). Unexpectedly, participants who only had to prepare a lesson, reported higher feelings of social presence compared to the participants who taught on video. Only the teaching on video participants reported higher levels of arousal compared to participants in the re-study control condition. There were no differences between the conditions on transfer performance.

The fourth study, which is presented in **Chapter 5**, was a conceptual replication study of the third study (Chapter 4) with the teaching on video, preparing to teach, and re-study control conditions. Again, the main research question was whether learning by teaching would help student teachers to mitigate the confirmation bias more than preparing to teach, or re-study. This time, the learning outcome was the open-mindedness displayed in a citizenship education lesson for primary school children aged 10-12 years. Furthermore, participants self-reported open-mindedness was measured one week before and immediately after the intervention. Also, social presence and arousal were measured with the same questionnaires as in the third study. In addition, conceptual knowledge of the instructional content was measured. The instructional content was aimed at designing open-minded citizenship education lessons by using a "consider-the-opposite" thinking strategy, i.e., by learning to reduce the confirmation bias. As a posttest, all participants designed an open-minded citizenship education lesson plan. Results showed that the mean scores on open-mindedness were higher on the posttest compared to the pretest. The mean accuracy and completeness of the lesson plan in the control condition was higher than in the teaching on video and preparing to teach conditions. An explorative, overall qualitative analysis revealed no differences with regard to the quality of the open-minded lesson plans between the three conditions. Feelings of social presence and arousal levels did not differ between the teaching on video and preparing to teach conditions.

Discussion of the main findings and directions for future research

In this chapter, we focus on overarching issues that contribute to the theory, namely the corroboration of learning effects of a "consider-the-opposite" instruction, the role of feedback on the reduction of the confirmation bias and applying generative learning strategies to support learning to reduce the confirmation bias.

A well-established finding from critical thinking (hereafter, abbreviated as CT) research, is that providing students with explicit instruction supports learning and enhances performance on various types of CT tasks (Abrami et al., 2015; Abrami et al., 2008; Bangert-Drowns & Bankert, 1990; Niu et al., 2013). This also applies to more specific debiasing interventions such as "consider-the-opposite" interventions, and other debiasing interventions aimed at learning unbiased reasoning (Adame, 2016; Beaulac & Kenyon, 2014; Correia, 2018; Dunbar et al., 2014; Dunbar et al., 2013; Heijltjes, 2014; Hirt & Markman, 1995; Korteling et al., 2021; Larrick, 2004; Lee et al., 2016; Lilienfeld et al., 2009; Lord et al., 1984; Morewedge et al., 2015; Mussweiler et al., 2000; Sellier et al., 2019; Van Peppen, 2020). In the studies presented in this dissertation, participants' confirmation bias was reduced after an instruction combined with practising confirmation bias tasks (Chapter 2), and feedback (see WE and CA conditions in Chapter 3, and all conditions in Chapter 4). Specific evidence regarding the necessity of instruction in our studies, was found in the study presented in Chapter 2 in which a no-treatment control condition was included. This group only completed a pretest and a posttest. Contrary to the four other conditions in that experiment who all received instruction, the participants in the no treatment control condition showed no significant performance increase from the pretest to the posttest, i.e., no evidence of a reduced confirmation bias. Hence, an instruction combined with practising confirmation bias tasks and preferably providing feedback on the tasks, is conditional to support learning to reduce confirmation bias.

Subsequently, we will discuss other relevant findings from the studies presented in Chapter 2 to 5, with regard to the effectiveness of various instructional design principles. In our studies - contrary to other findings in the literature – we found no evidence that domain-specificity, elaborative feedback, and learning by teaching on video, resulted in an additional confirmation bias reduction relative to a control condition (Abrami et al., 2015; Bangert-Drowns & Bankert, 1990; Bangert-Drowns et al., 1991; Butler et al., 2013; Ennis, 1989; Fiorella & Mayer, 2013, 2014; Heijltjes, Van Gog, & Paas, 2014; Hoogerheide et al., 2016; Hoogerheide et al., 2014b; Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019; Marin & Halpern, 2011; Niu et al., 2013; Shute, 2008).

Domain-specific instruction and topic interest

In general, instruction which features authentic tasks that are perceived as relevant by the learner and which is provided within the context of a specific subject, is assumed to have a positive effect on learning CT, and more specifically on reasoning and decision making (Abrami et al., 2015; Bangert-Drowns & Bankert, 1990; Brandsford et al., 1999; Brown et al., 1989; Dumitru, 2012; Hill, 2000; Marin & Halpern, 2011; McPeck, 1990; Moore, 2004; Niu et al., 2013; Nursa'adah et al., 2021; Sá et al., 1999; Tiruneh, De Cock, et al., 2018; Tiruneh, Gu, et al., 2018; Van der Zanden et al., 2018). For example, authentic tasks in higher education often have an explicit link to a future profession. The positive effect of authentic tasks on learning and performance might be due to the fact that authentic problems, perhaps due to their applied nature and explicit link to the future profession (i.e., domain-specificity), spark motivation, which in turn might trigger learning behaviour that enhances performance (Deci & Ryan, 1985; Duncan & Pintrich, 1992; Hidi, 2001; Mason et al., 2008; Schraw et al., 1995; Wijnia et al., 2014).

The relationship between motivation and performance is, however, not straightforward. Participants in the domain-specific (educational) condition of the first study presented in Chapter 2, reported higher topic interest than the participants in the general condition, suggesting that motivation was higher in the more authentic condition than in the general one. Yet, performance, expressed as confirmation bias reduction, did not differ between the two conditions. This observed lack of correlation between an indicator of motivation and performance seems to be consistent with other findings in the literature. For example, Garon-Carrier et al. (2016) showed that motivation was not causally related to performance. In fact, the reversed appeared to be true in their experiment: Primary school pupils' mathematics performance had a significant effect on their motivation, but this relationship was not reciprocal: motivation had no effect on performance. In addition, McConney et al. (2014) even found a negative relationship between motivation and performance: Science students who were highly committed to inquiry-oriented learning, showed below-average levels of science literacy and above-average levels of interest and engagement in science in McConney et al.'s study (2014).

A possible explanation of the findings in Chapter 2 might be related to a finding from the systematic review on CT practices in teacher education by Lorencová et al. (2019). They found that motivation was only positively related to performance in case of longer instructional interventions and that a short intervention is regarded as a constraining factor. In our study on domain-specifity, the intervention might have been too short to show positive effects of motivation on performance. A suggestion for future research, therefore, is to extend the instruction. A performance difference might emerge in a longer instructional sequence. In such a sequence, the difference in sustained attention and time on task between a domain-specific instruction and a general instruction will likely increase in favour of the former condition, which in turn might result in a performance advantage. Yet, interest driven sustained attention and extra time on task might not be sufficient to enhance performance. It might additionally be required to provide feedback, aimed at helping students to monitor and improve their conceptual understanding of the topic at hand. Investigating this causal path specifically for confirmation bias might be an interesting avenue for future research.

Elaborative feedback

Elaborative feedback, in which participants not only receive the correct answer to a task but also feedback on the solution steps that lead up to the correct answer, has produced well-established learning gains in various learning domains (Bangert-Drowns et al., 1991; Bown, 2017; Butler et al., 2013; Candel et al., 2021; Finn et al., 2018; Hattie & Timperley, 2007; Olms et al., 2017; Shute, 2008; Van der Kleij et al., 2012; Van der Kleij et al., 2015). The beneficial learning effects of elaborative feedback can be explained by a particular effect from the Cognitive Load Theory (Paas, Renkl, et al., 2003; Sweller, 1988; Sweller et al., 1998, 2019), namely the worked example effect (Schwonke et al., 2009; Sweller, 1988, 2006; Van Gog et al., 2011). This effect entails that studying worked examples during instruction produces better learning with less investment of time and mental effort than solving conventional problems because worked examples support learners' construction of cognitive problem-solving schemas, which can regulate their learning during practice (Cooper & Sweller, 1987; Paas, 1992; Sweller et al., 1998, 2019; Van Gog & Rummel, 2010; Ward & Sweller, 1990). In our study presented in Chapter 3, however, no differences were found between a condition that received elaborative feedback in the form of worked examples on confirmation bias practice tasks (the WE condition) and a condition that received only the correct answer as feedback on confirmation bias practice tasks (the CA condition).

Based on the feedback literature (Hattie & Timperley, 2007; Roelle et al., 2017; Van der Kleij et al., 2015; Vollmeyer & Rheinberg, 2005), there might be several reasons why elaborative feedback was not more effective in our study than correct answer feedback. A possibility is that participants processed the elaborative feedback in the worked example condition but did not understand the provided explanation. This in turn might have left them with the correct answer and – potentially – with a rule of thumb for solving these kinds of tasks. In this situation, the lack of understanding the feedback would explain our results.

It might also be possible that participants did not process the elaborative feedback provided in the worked example. That might be due to several reasons related to motivation or lack of perceived necessity. In any case, it would result in a lack of difference between the two feedback conditions. This latter explanation is somewhat supported by our data. In the study the two conditions did not differ in time on task. This could mean that the elaborative feedback was superficially processed. However, it could also indicate that students in the correct answer condition thought about the rationale underlying the correct answer and that this reduced the effect of elaboration in the worked example condition. Another possibility is that participants processed the elaborative feedback in the worked example condition, understood the elaboration but might not have needed it to improve their performance. In this case, feedback is not effective, and participants might have ignored the feedback message (Hattie & Timperley, 2007; Roelle et al., 2017; Van der Kleij et al., 2015). Perhaps, the correct answer (which was provided in both conditions) made it possible for participants to derive a heuristic, which was sufficient to lead to performance gains on the types of tasks we offered them. For example, participants might have deduced from the correct answer on the Wason's four-card selection task that they should turn the card, which confirms the rule and the card that invalidates the rule. Using this rule of thumb will enhance performance thereby limiting (or even erasing) the potential benefit of additional understanding, which might have resulted from processing the elaborative feedback. In sum, at the moment, it is only possibly for us to speculate about why the two feedback conditions did not differ in performance. To unearth the mechanism(s) underlying the lack of effect, future studies might include measures that can be used to gain more insight in how students process the feedback both qualitatively and quantitatively.

Generative learning

Generative learning strategies are effective for knowledge building and research shows robust positive effects on memory and transfer of knowledge, compared to often used but rather superficial learning strategies, i.e., re-study, summarizing, and highlighting (Blasiman et al., 2017; Chi, 2009; Dunlosky et al., 2013; Fiorella & Mayer, 2016). Generative learning is "a sense-making activity, which involves building cognitive structures that can be used in new situations" (Fiorella & Mayer, 2016, p. 733). Examples of effective generative learning strategies are practice testing, interleaved practice, distributed practice, self-explaining, and learning by teaching (Dunlosky et al., 2013; Fiorella & Mayer, 2016). A specific and fairly new learning by teaching strategy is learning by teaching on video (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2016; Hoogerheide et al., 2014b; Hoogerheide, Renkl, et al., 2019; Hoogerheide, Visee, et al., 2019; Kobayashi, 2019). The benefits of learning by teaching on video are attributed to raising feelings of social presence which in turn, evoke arousal, i.e., the social presence hypothesis. The social presence hypothesis assumes that the awareness of a fictitious audience during teaching induces motivational processes that support learning and more transfer in an oral teaching on video condition than in a written preparing to teach condition (Biocca & Harms, 2002, October; Cui, 2013; Gunawardena, 1995; Gunawardena & Zittle, 1997; Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019; Lowenthal, 2012; Oh et al., 2018; Slater et al., 1999; Tu, 2002; Weidlich & Bastiaens, 2017). However, in the study with a teaching on video condition and a re-study control condition by Hoogerheide, Renkl et al. (2019), it was shown that situations that are higher in social presence induce more arousal than situations lower in social presence but that performance gains cannot be attributed to arousal, but more likely to the cognitive processes that are induced by considering a potential audience.

In the two studies presented in Chapter 4 and 5, no evidence was found for the social presence hypothesis. The condition with the assumed highest social presence, teaching on video, did not differ on mean arousal scores compared to only preparing to teach. Feelings of social presence were highest in the preparing to teach condition in the experiment presented in Chapter 4 and did not differ between both conditions in the experiment presented in Chapter 5. In addition, no significant differences between teaching on video and preparing to teach on learning and transfer were found in the third experiment (Chapter 4). Contrary to our expectations, participants in the control condition prepared a better lesson plan in the study presented in Chapter 5 compared to both generative learning conditions. Giesbers et al. (2014), found similar results, i.e., no improved performance in a long-termed online distance programme in economics, when learning through technologies that are associated with higher degree of social presence (e.g., web video conferencing) compared to lower degree of social presence (e.g., discussion forums).

The fact that we did not find differences in feelings of social presence between the teaching on video condition and the preparing to teach condition in the studies from this dissertation might be due to the way in which we measured social presence. This was done through self-reports after the intervention. Feelings of social presence might have faded away because the actual act of teaching was over. Other measures, such as wristbands, that measure a direct change in the electrical resistance or temperature of the skin caused by e.g., arousal (e.g., Biocca & Harms, 2002, October; Cui, 2013; Gunawardena & Zittle, 1997; Hoogerheide, Renkl, et al., 2019), or counting the number of personal references in the explanation (e.g., Hoogerheide et al., 2016; Jacob et al., 2020; Lachner et al., 2018) might have yielded other results.

Furthermore, in our studies, performance did not differ between the teaching on video condition and the preparing to teach condition. Participants in both conditions had to prepare their lesson on paper. Afterwards, the teaching participants had to read their explanation aloud to the camera of their own smartphone or laptop. This might have resulted in knowledge reproduction rather that the intended knowledge building, which would have had little additional impact on the final test performance. Another explanation of the absence of performance differences might have been the quality of the explanations. This quality was not particularly high in both conditions in both experiments. Effects of self-explanation on learning and performance are contingent on a sufficiently high quality of self-explanations. It might be possible, that the level of understanding participants reached after instruction and practice tasks was not high enough to allow for beneficial effect of teaching on video and preparing to teach to emerge (e.g., Jacob et al., 2021). All in all, it is still not fully clear when increased social presence is desirable or not for learning (Lachner et al., 2021), especially in the context of generative learning for novices (Jacob et al., 2021). Further research is needed to examine when and why feelings of social presence are beneficial for learning, and in which learning contexts.

Methodological issues

A common methodological issue in heuristics and biases research is the reliability of the measurements. For the three posttests of the first study, the posttest and transfer-test of the second study, and the posttest of the third study, the Cronbach's alphas were low considering the use of test scores to compare conditions ($\alpha < .60$). This might be due to the absence of a general factor of proneness to solve cognitive bias tasks (Berthet, 2021; Bruine de Bruin et al., 2007; Janssen, Mainhard, et al., 2019; Janssen et al., 2020; Korteling et al., 2021). We used two types of tasks in the first three studies, that are frequently used in confirmation bias research (e.g., Evans et al., 1993; Heijltjes, Van Gog, Leppink, et al., 2014; Heijltjes, Van Gog, & Paas, 2014; Jonas et al., 2001; Sellier et al., 2019; Stanovich & West, 2008; Van Peppen et al., 2018; Wason, 1968). For the hypothesis-testing task based on Jonas et al. (2001), participants had to decide which information is needed to take a reasoned stand. To solve Wason's four-card selection task, one had to have knowledge about and apply the formal rules of logic. Participants probably did gain the right mindware through instruction and practice tasks with feedback to solve the types of tasks based on Jonas et al (2001), but not for the not instructed Wason's four-card selection task. Therefore, the correlation between the knowledge and skills of these two types was low, which would have had a negative impact on Cronbach's alpha. Another explanation for a somewhat lower Cronbach's alpha is a possible ceiling effect regarding solving the hypothesis-testing Jonas-task. For this task, participants might have applied a heuristic that led to the right answer. They found out that the principle for solving this type of task was to choose balanced information. This might have been an easy task to solve, and participants might have applied Type 1 thinking instead of the desired Type 2 thinking. A ceiling effect has a negative impact on the correlations between items: Without variance on 1 item, by definition no correlation is possible with other items. This lowers Cronbach's alpha.

When the reliability is low, it is hard to detect small effects because a relatively large part of the distribution of the scores in non-systematic (error variance). However, large effects can still be found, and we might assume that the effects we found were large enough to be picked up in the statistical tests. Future studies could increase the power of statistical tests by using items of one type of task and trying ty avoid ceiling (and bottom) effects by modifying the current instruction and practice tasks.

Another issue is related to the validity of the measurement instruments that we used. As discussed above, our participants might have applied rules of thumb and therefore the measurement scores in our studies might not only reflect the level of confirmation bias. For future research, it might, therefore, be useful to adapt the tasks that we used in our measurement instruments. For example, by asking participants to explain how they solved the hypothesis-testing tasks and asking them to substantiate their initial opinion. In addition, participants have to explain whether or not, and why they changed their opinion afterwards. This way, these kinds of tasks might provide an even stronger indication of confirmation bias than the tasks that we employed in the studies of this dissertation. Likewise, we asked participants to choose four out of eight summaries of bogus articles that were offered by us, but for future research it might be interesting to let participants search for articles themselves. In the last study, described in Chapter 5, we measured confirmation bias in a more ecological valid manner for student teachers, namely with the preparation of lesson plans. This task might be extended with observing the actual teaching – and the degree to which confirmation bias occurs - in the classroom.

Open Science

A strength of the research presented in this dissertation, is that three studies were pre-registered and, hence, made the hypotheses, materials, planned data-collections, planned methods, planned data-analyses, and the data sets of the studies freely accessible and available to anyone on the online repository "Open Science Framework" (OSF). By using open science practice, researchers make important decisions early on and these decisions are transparent. This contributes to reduce biases that occur once the data are collected or published. Pre-registration supports clearly separating exploratory from confirmatory research: It will increase the distinct between hypothesis-generating and hypothesis-testing and subsequently increase the interpretability, reproducibility, and credibility of research findings (Nosek et al., 2018). Open practices are important to reduce some questionable but sometimes unintentionally research practices such as HARKing (Hypothesizing After the Results are Known, Kerr, 1998; Simmons et al., 2011), or data dredging (also known as p-hacking, searching the data for patterns that are statistically significant, Munafò et al., 2017; Nosek & Lakens, 2014). It will decrease false positives in research, because open science restricts researcher's degrees of freedom (Nosek et al., 2018; Wicherts et al., 2016). Open science can help to turn "the replicability crises into a credibility revolution" (Plucker & Makel, 2021, p. 91), strengthen confidence in scientific findings, and advance educational psychology research faster and more efficiently because knowledge about what works and does not work (yet) is readily available.

Practical implications

Based on the studies described in this dissertation, practical implications can be identified for educational practice. From our findings, it is clear that in a highly controlled setting, instruction about the confirmation bias and how to reduce this bias and practice tasks with feedback has a positive effect on reducing the confirmation bias in hypothesis-testing tasks. In a real classroom setting, paying attention to the confirmation bias and how to reduce it would therefore be advised. In such settings, however, the didactic approach would be somewhat different from the ones used in our studies. A suggestion for educational practice is provided by Abrami et al. (2015). They suggest the instruction should contain a triple combination of dialogue, mentoring and authentic instruction (i.e., Three Category Scheme), because these three factors work as a catalyst for CT learning. Methods are for example, class debates and discussions, Socratic dialogues, role playing, case studies, simulations, games, and internships. Classroom studies that featured all three types of interventions led to significant larger effect sizes compared to interventions that only featured one type of intervention as was the case in the experiment described in this dissertation (good practices of classroom studies on CT can be found in studies by e.g., Bell & Loon, 2015; Dumitru, 2012; Hill, 2000; Tiruneh et al., 2016; Van Peppen, 2020; Zohar & Schwartzer, 2005).

Participants in the four experiments of this dissertation were mainly student teachers. The instruction and practice tasks supported their learning to reduce the confirmation bias in the controlled settings and with the confirmation bias tasks we used. Within their future practice as a primary school teacher, it is important that they are aware of the negative impact of the confirmation bias and that it hampers good judgment and decision making. Teachers might have biased expectations of their pupils, or a teacher's confirmation bias may result in one-sidedness where open-mindedness and multiple perspectives should be considered (Darley & Gross, 1983; De Boer et al., 2018; Rosenthal & Jacobson, 1968; Timmermans et al., 2018). Hence, it is important that (student) teachers learn and practice to reduce their confirmation bias in real settings and with other tasks as well. Learning about the confirmation bias and learning to take perspective by considering alternatives and opposites and learning to become more open-minded, might be a part of the primary teacher education curriculum. Therefore, it is important that the teacher educators themselves are equipped with the skills and knowledge and that they are positive towards the relevance to teach unbiased reasoning and CT skills such as perspective taking (Janssen, Mainhard, et al., 2019; Janssen, Meulendijks, et al., 2019). Janssen et al. (2019) concluded from their research on teaching CT in higher education, that educators themselves must be trained to be able to avoid bias in their own reasoning, to recognize students' biased reasoning, to provide feedback to their students why their reasoning is biased, and to have a positive disposition towards teaching CT, i.e., acknowledge the importance of CT in education.

Next to educators, curriculum developers and advisors are advised to embed learning to reduce the confirmation bias into the curriculum. To support educators in the educational practice when designing curricula in which CT skills are explicitly considered, professional development courses are recommended (Janssen, Mainhard, et al., 2019), and sharing educational and scientific sources and good practices to develop and teach their own courses can play an important role (e.g., the Dutch online platform Kritisch Leren Denken: https://www.kritischdenkenhbo.nl/).

All of this assumes that, in terms of best practice activity to improve learning to reduce the confirmation bias, there are promising instructional design principles, but more research is needed to explore under which conditions these are most effective for learning. Instruction in a domain-specific learning context, or the use of the generative learning strategy "learning by teaching on video" have the potential to improve learning, but more research is needed to acquire more insight into the beneficial mechanisms underlying these strategies.

Educational models and the content of CT in the curriculum have been proposed as part of the "21st century skills", but not much is known yet what teacher knowledge is conditional and how to effectively teach CT in primary school (Ab Kadir, 2017; Nickerson, 1988). Additionally, based on the 21st century skills ideas, (student) teachers have the responsibility to turn their pupils in primary school into good critical thinkers. Therefore, they need to be convinced about the importance of fostering young children's CT and to be confident about their own CT abilities (Lorencová et al., 2019). The merits of CT are undoubtedly for young children as well: For their personal development and being an involved citizen. CT is an essential skill for examining and assessing information, determining positions and making informed decisions. CT presupposes analytical thinking and an open, inquiring attitude. Focusing on reducing the confirmation bias as a part of CT, one may assume that the teacher has a crucial role in effectively teaching "consider-the-opposite" in primary school and guide children to become critical and unbiased thinkers. Therefore, we suggest intertwining this "double loop" of learning ánd learning to teach CT, as goals of the educational CT models in research with a focus on effective instructional design principles for primary school to further investigate how to effectively teach young children the essential CT skills and dispositions. In sum, every teacher must be trained and feel competent enough to teach CT skills.

Nederlandse samenvatting (Summary in Dutch)

Ontwerpprincipes voor het onderwijs om de voorkeur voor bevestiging te verminderen

Inleiding

Wie openstaat voor de wereld, kan allerlei mogelijkheden benutten om zich breed te informeren over verschillende standpunten. Maar dat gebeurt niet altijd. Op het internet wordt bijvoorbeeld gebruik gemaakt van algoritmen: de informatie die we te zien krijgen, is aangepast aan de zoektermen die meestal gericht zijn op de bevestiging van onze standpunten waardoor informatie over andere opvattingen niet wordt getoond (Quatrociocchi et al., 2016). Ons beeld wordt gefilterd en dit kan leiden tot eenzijdige opvattingen en bevooroordeeld redeneren (Vydiswaran et al., 2015; Zimmer et al., 2019).

Toch gebeurt bevooroordeeld redeneren niet pas nu en niet alleen digitaal. Fouten in het menselijk redeneren zijn van alle tijden omdat het samenhangt met de manier waarop mensen redeneren. Om onze redeneerprocessen te vergemakkelijken, vertrouwen wij namelijk op heuristieken (Tversky & Kahneman, 1974). Heuristieken zijn vuistregels op basis van eerdere ervaringen en opgebouwde kennis. Ze zorgen ervoor dat we beslissingen relatief eenvoudig en snel kunnen nemen. Dit is vooral handig in situaties waarin niet alle benodigde informatie beschikbaar is. Heuristieken leiden ons dan snel naar mogelijke oplossingen en besparen ons tijd. Ook kunnen heuristieken ons werkgeheugen ontlasten. Ze verminderen dan de cognitieve inspanning die nodig is om beslissingen te nemen en oordelen te vellen (Stanovich, 2011; Tversky & Kahneman, 1974). Heuristieken kunnen echter ook systematische denkfouten opleveren, de zogenoemde biases (Tversky & Kahneman, 1974).

Het ontstaan van systematische denkfouten kan worden verklaard door twee soorten redeneerprocessen: Type 1 en Type 2 denken, ook wel Systeem 1 en Systeem 2 denken genaamd (De Neys, 2021; Evans, 2003; Evans & Stanovich, 2013; Kahneman, 2011; Stanovich & West, 2008). Type 1 denken verloopt automatisch, is gebaseerd op onze heuristieken, vergt weinig inspanning en is efficiënt en effectief in veel routinematige situaties (Evans, 2003). Type 1 denken is erg handig in het dagelijks leven, bijvoorbeeld bij het oplossen van een som zoals 2 x 2 of bij het lopen van de route naar de dichtstbijzijnde bakkerij. Type 2 denken daarentegen verwijst naar het meer bewuste, reflectieve denksysteem dat betrekking heeft op rationaliteit en sequentieel denken. In situaties waarin iets op het spel staat en wanneer informatie kritisch moet worden beoordeeld, moeten mensen geen automatische, maar juist **reflectieve** Type 2 denkprocessen gebruiken. Wanneer weloverwogen oordelen of belangrijke beslissingen vereist zijn, schakelen mensen ideaal genomen dan het Type 2 denken in en onderdrukken ze het Type 1 denken (Kahneman, 2011). Voorbeelden van situaties waarin Type 2 denken gebruikt moet worden zijn het bepalen van het middelbare schooladvies in groep 8 van de basisschool en beslissen op welke politieke partij men gaat stemmen.

De dual process theorie van Evans (2003) geeft aan dat overmatig vertrouwen op Type 1 denken waar Type 2 denken passender zou zijn, kan leiden tot denkfouten (biases). Mensen realiseren zich dan niet dat een Type 2 denken nodig is, of ze passen geen Type 2 denken toe omdat het te veel moeite kost, of ze hebben niet de juiste kennis en vaardigheden om een passende Type 2 denken op te wekken.

Er zijn verschillende soorten biases. Een veelvoorkomende bias is de voorkeur voor bevestiging (de zogenoemde confirmation bias, Nickerson, 1998). Mensen hebben de neiging om meer aandacht te schenken en meer waarde te hechten aan informatie die hun ideeën en overtuigingen bevestigt, dan aan informatie die deze ideeën en overtuigingen tegenspreekt (Nickerson, 1998). Zo zoeken ze bevestigend bewijs voor hun eigen standpunt bij het evalueren van concurrerende standpunten, interpreteren ze bewijs verkeerd om hun overtuigingen te ondersteunen, en hechten ze meer belang aan bevestigend dan aan ontkrachtend bewijs wanneer ze nadenken over hun overtuigingen (Charness & Dave, 2017; Nickerson, 1998). Deze bias is gebaseerd op cognitieve processen die onbevooroordeeld Type 2 redeneren verstoren omdat ze bijvoorbeeld een te groot beroep doen op het werkgeheugen. Mensen verkiezen dan Type 1 denken.

De confirmation bias kan in de praktijk tot problemen leiden, bijvoorbeeld bij het diagnosticeren door een arts, het argumenteren in de wetenschap en bij het doen van gerechtelijke uitspraken (Nickerson, 1998, pp. 192-194). Ook in de context van het basisonderwijs kan de confirmation bias verstrekkende gevolgen hebben. Stereotypen en deels onjuiste overtuigingen over de sociaaleconomische achtergrond of de intelligentie van leerlingen, kunnen resulteren in vooringenomen verwachtingen van leraren die negatief kunnen uitpakken voor de leerlingen (De Boer et al., 2018; Kelley, 1950; Timmermans et al., 2018). De confirmation bias kan dan in het onderwijs van invloed zijn op bijvoorbeeld de resultaten van leerlingen en het middelbare schooladvies (self fulfilling prophecy; Darley & Gross, 1983; Foster et al., 1976; Rosenthal & Jacobson, 1968; Timmermans et al., 2015; Van den Bergh et al., 2010).

Het effect van de confirmation bias wordt door het gebruik van algoritmen nog eens versterkt. Quatrociocchi et al. (2016) toonden aan dat Facebook-gebruikers meestal interacties hebben met gelijkgestemden die hen voeden met bevestigende informatie over hun standpunt en weerleggende informatie afwijzen. Bovendien wordt de confirmation bias versterkt door de "filterbubbel". Dit betekent dat het algoritme van een zoekmachine alleen informatie weergeeft waar de internetgebruiker het waarschijnlijk mee eens is en tegengestelde standpunten weg filtert.

Om de confirmation bias te verminderen en te leren om onbevooroordeeld te redeneren, zouden mensen gebruik moeten maken van tegengestelde informatie. Door het aanbieden van alternatieve informatie tijdens een beoordelings- of beslissingsproces – waardoor perspectiefwisseling aangemoedigd wordt – kan de confirmation bias in uiteenlopende situaties effectief tegengegaan worden (een strategie die bekend staat onder de naam "Consider the opposite"; Hirt & Markman, 1995; Koriat et al., 1980; Lord et al., 1984; Lord et al., 1979; Morewedge et al., 2015; Snyder & Swann, 1978; Soll et al., 2015). Deze strategie moedigt mensen actief aan om hun aandacht te richten op alternatieve zienswijzen die anders niet in overweging zouden worden genomen (Danielson & Sinatra, 2017; Larrick, 2004). Hierdoor worden waarschijnlijk Type 2 denkprocessen geactiveerd die resulteren in beter onderbouwde besluiten en oordelen die zijn gebaseerd op meer perspectieven.

Eerdere studies naar de vermindering van de confirmation bias hebben perspectiefwisseling gebruikt om theoretische mechanismen te toetsen die ten grondslag liggen aan de confirmation bias (Hirt & Markman, 1995; Lord et al., 1984; Lord et al., 1979). Deze inzichten zijn ook nuttig voor het onderwijs. Voor het onderwijs geldt daarnaast de specifieke vraag hoe je instructie en oefening zodanig kunt inrichten, dat leerlingen zelf leren om de confirmation bias te verminderen. Binnen onderzoek naar de confirmation bias is hiernaar nog weinig onderzoek gedaan maar er zijn wel belangrijke inzichten over het instructieontwerp opgedaan bij andere cognitieve biases. Studies naar het verminderen van cognitieve biases in het algemeen hebben aangetoond dat er een debiasing effect optreedt wanneer deelnemers (1) expliciete instructie krijgen over de kennis die nodig is om onbevooroordeeld te redeneren en het belang ervan inzien, (2) de mogelijkheid krijgen om te oefenen met relevante (domeinspecifieke) redeneertaken en (3) feedback krijgen op de oefentaken (Adame, 2016; Dunbar et al., 2014; Heijltjes, Van Gog, Leppink, et al., 2014; Mussweiler et al., 2000; Sellier et al., 2019; Van Peppen, 2020). Er zijn echter nog allerlei vragen over hoe je een instructie zo kunt ontwerpen dat de confirmation bias wordt verminderd. De vraag die in dit proefschrift centraal staat, was dan ook hoe een instructie om de confirmation bias te verminderen, het beste kan worden vormgegeven.

In de studies van dit proefschrift zijn een aantal factoren gemanipuleerd om antwoord te krijgen op deze vraag. Wanneer een instructie wordt gegeven in de context van een specifiek onderwerp, is er sprake van **domeinspecifiek leren**. Als het leren van een algemene vaardigheid of algemene leerstofinhoud het doel is van de instructie, dan spreken we van een algemene benadering (Abrami et al., 2008; Ennis, 1989). Domeinspecifiek leren heeft echter in vergelijking met de algemene benadering een positief effect op bijvoorbeeld redeneren en beslissingen nemen (Bangert-Drowns & Bankert, 1990; Heijltjes, Van Gog, & Paas, 2014; Marin & Halpern, 2011; Niu et al., 2013). Leren in een domeinspecifieke context kan de interesse in het onderwerp en de daaropvolgende prestaties verhogen vanwege de toegepaste aard en de expliciete link met het toekomstige beroep. Een grotere interesse in een onderwerp leidt tot een hogere motivatie en meer tijd besteed aan zelfstudie en hoe meer tijd aan zelfstudie wordt besteed, hoe beter de prestaties - tenminste als de zelfstudietijd op een effectieve manier wordt ingevuld (Deci & Ryan, 1985; Duncan & Pintrich, 1992; Hidi, 2001; Mason et al., 2008; Schraw et al., 1995; Wijnia et al., 2014). De relatie tussen motivatie en prestatie is echter niet eenduidig (zie bijvoorbeeld de studies van Garon-Carrier et al., 2016; McConney et al., 2014). De mate van domeinspecificiteit van de leercontext zou dus een positieve invloed kunnen hebben op bijvoorbeeld het leren van leraren in opleiding als de instructie en de oefentaken gekoppeld worden aan lesgeven als beroep zoals het voorbereiden en uitvoeren van een onderwijsactiviteit. Om dit te testen hebben we in de studie die beschreven staat in hoofdstuk 2 de leercontext gemanipuleerd.

Het geven van **feedback** ondersteunt het leren ook. Feedback kan worden gedefinieerd als de informatie die een leraar aan een student geeft over zijn/haar prestaties of begrip met als doel het leren en presteren te bevorderen (Hattie & Timperley, 2007). Het was nog niet bekend welke rol feedback kan spelen bij het leren verminderen van de confirmation bias, maar het is een krachtig leermiddel dat bijdraagt aan betere prestaties, begrip en toepassing van de leerstof (Hattie, 2009, 2012; Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2005). Uitgewerkte voorbeelden kunnen worden beschouwd als een vorm van uitgebreide feedback. Uitgebreide feedback is effectiever voor het leren dan alleen het geven van het juiste antwoord als feedback (Butler et al., 2013). Uitgebreide feedback is bijvoorbeeld een stapsgewijze uitleg waarin wordt voorgedaan hoe iemand tot de juiste oplossing kan komen (Hattie & Timperley, 2007; Shute, 2008; Sweller et al., 1998; Van der Kleij et al., 2015). Het leereffect van uitgebreide feedback in de vorm van een stapsgewijze uitleg hoe je tot het antwoord moet komen, houdt verband met het worked example effect dat kan worden verklaard vanuit de Cognitive Load Theory (Paas, Renkl, et al., 2003; Sweller, 1988; Sweller et al., 1998, 2019). Deze theorie geeft aan dat het bestuderen van uitgewerkte voorbeelden tijdens de instructie leidt tot beter leren in minder tijd en minder mentale inspanning in vergelijking met het leren door het oplossen van conventionele problemen. Uitgewerkte voorbeelden leiden tot de constructie van cognitieve schema's over de aanpak om problemen op te lossen (Cooper & Sweller, 1987; Paas, 1992; Sweller & Cooper, 1985; Sweller et al., 1998; Van Gog & Rummel, 2010; Ward & Sweller, 1990). Conventionele problemen zijn problemen waarvoor studenten zelf de oplossing moeten vinden en die vaak veel mentale inspanning vergen voor het oplosproces wat ten koste gaat van leren (Cooper & Sweller, 1987). In de tweede studie, die wordt beschreven in hoofdstuk 3 van dit proefschrift, is getest wat het effect is van een uitgewerkt voorbeeld als feedback, feedback met het juiste antwoord en geen feedback is op het verminderen van de confirmation bias.

Naast de besproken factoren aan de aanbodkant bij het leren, is een belangrijke factor die effectief leren beïnvloedt hoe studenten de informatie verwerken (Fiorella & Mayer, 2016). Als studenten leren, maken ze vaak gebruik van suboptimale leerstrategieën die leiden tot oppervlakkig leren, zoals samenvatten of herlezen (Blasiman et al., 2017; Dirkx et al., 2019; Dunlosky et al., 2013; Karpicke et al., 2009). De meest effectieve leerstrategieën zijn echter generatieve verwerkingsstrategieën (generative learning strategies). Generatieve verwerking ondersteunt het leren omdat het een constructieve handeling is. Hierbij spannen studenten zich in om betekenis te geven aan de leerstof door nieuwe ervaringen en kennis te integreren met hun bestaande kennisstructuren of schema's (Fiorella & Mayer, 2016). Hierdoor wordt de leerstof cognitief geherstructureerd en dat leidt tot de koppeling van de nieuwe leerstof met voorkennis (Chi, 2009; Fiorella & Mayer, 2016). Een voorbeeld van een generatieve verwerkingsstrategie is zelfverklaren (self-explanation; Chi et al., 1989; Dunlosky et al., 2013; Rittle-Johnson et al., 2017). Zelfverklaren is effectief omdat door het uitleggen van nieuwe informatie het begrip wordt bevorderd door de integratie van nieuwe kennis en voorkennis. Ook maakt het studenten meer bewust van hun redenering naar de oplossing en van nog bestaande kennishiaten die hersteld moeten worden (Chi et al., 1989; Dunlosky et al., 2013; Fiorella & Mayer, 2016; Rittle-Johnson et al., 2017).

Een variant van zelfverklaren is leren door les te geven via een video aan een fictief publiek (learning by teaching on video; Fiorella & Mayer, 2013, 2014, 2016, 2018; Fiorella et al., 2020; Kobayashi, 2019). Een student profiteert zelf van het uitleggen aan anderen, omdat hij/zij eerst de relevante informatie selecteert om in de uitleg op te nemen, deze daarna zodanig organiseert dat deze door anderen kan worden begrepen, en ten slotte door het materiaal uit te werken en te combineren met bestaande kennis, wat leidt tot betekenisvolle kennisstructuren (Duran, 2017; Fiorella & Mayer, 2016).

De gunstige leereffecten van leren door les te geven via video kunnen daarnaast worden verklaard door de social presence hypothese. Deze hypothese gaat ervan uit dat het bewustzijn van de aanwezigheid van anderen in een (virtuele) leeromgeving leidt tot betere leerprestaties (Biocca & Harms, 2002, October; Gunawardena, 1995; Lowenthal, 2012; Slater et al., 1999). Gevoelens van social presence wekken arousal op: een toestand waarin iemand door spanning geactiveerd wordt met als resultaat een toenemende hartslag, bloeddruk en transpiratie. Zowel social presence als arousal leiden ertoe dat iemand wordt gestimuleerd om zo nauwkeurig en volledig mogelijk de leerstof uit te leggen, zodat het publiek de inhoud begrijpt. De eerdergenoemde voordelen van zelfverklaren gaan hier dus ook op. Het bewustzijn van de aanwezigheid van fictieve anderen kan resulteren in beter leren en een betere transfer van leren naar een andere context, in vergelijking met bijvoorbeeld herlezen vanaf papier of het schriftelijk samenvatten van de leerstof. Recente onderzoeken hebben veelbelovende resultaten laten zien van het lesgeven aan een fictief publiek in vergelijking met alleen het voorbereiden van een les (preparing to teach, expectation to teach) (Hoogerheide et al., 2016; Kobayashi, 2019). De effecten op leren door les te geven via video zijn echter nog niet onderzocht in de context van het leren verminderen van de confirmation bias. Dit stond centraal in de studies die beschreven staan in hoofdstuk 4 en 5 van dit proefschrift.

De samenvatting van de afzonderlijke onderzoeken

In dit proefschrift werden de effecten van de bovengenoemde factoren, d.w.z. de domeinspecificiteit van de leercontext (hoofdstuk 2), feedback (hoofdstuk 3) en de generatieve verwerkingsstrategie teaching on video (hoofdstuk 4 en 5), op de vermindering van de confirmation bias onderzocht in een sequentie van instructie en oefening. De instructie bestond uit een uitleg over wat de confirmation bias is, waarom het belangrijk is om deze te leren verminderen en een uitleg over de denkstrategie consider the opposite. Oefening vond plaats aan de hand van taken die veelvuldig gebruikt worden om de confirmation bias te meten (Jonas et al., 2001; Wason, 1968). Het gaat om taken waarin het testen van een hypothese centraal staat gebaseerd op de studie van Jonas et al. (2001) en selectietaken met vier kaarten (Wason, 1968). Hieronder staat een voorbeeld van achtereenvolgens de hypothese test taak gebaseerd op Jonas et al. (2011) en een voorbeeld van een selectietaak met vier kaarten (Wason, 1968), beide uit de domeinspecifieke context.

Een voorbeeld van een hypothese test taak gebaseerd op Jonas et al. (2011):

A) Lees de volgende stelling:

Kinderen van nu hebben minder doorzettingsvermogen dan kinderen van vroeger.

Ben je het eens of oneens met de stelling? Geef je standpunt aan, gebaseerd op je eerste ingeving.

- o Eens
- o Oneens

Er is meer informatie beschikbaar over de stelling **Kinderen van nu hebben minder doorzettingsvermogen dan kinderen van vroeger**. Hieronder kun je lezen welke informatie beschikbaar is om je standpunt beter te onderbouwen. Je kunt ervan uitgaan dat alle informatie even betrouwbaar is. Lees de informatie eerst aandacht door en kies vervolgens op basis van de samenvatting vier artikelen die je over dit onderwerp zou willen lezen.

- ☐ 1. In dit artikel stellen de auteurs dat kinderen geen doorzettingsvermogen meer hebben vanwege de moderne opvoeding: pedagogische weifelachtigheid en gemakzuchtige verwennerij hebben een generatie van onuitstaanbare prinsjes en prinsesjes gebaard.
- ☐ 2. In dit artikel stellen de auteurs dat kinderen altijd doorzettingsvermogen hebben omdat ze de intrinsieke wil hebben om te leren en zich te ontwikkelen.
- ☐ 3. In dit artikel stellen de auteurs dat kinderen geen doorzettingsvermogen meer hebben omdat ouders door hun eigen drukke leven steeds toegeven aan alles wat hun kinderen vragen.
- ☐ 4. In dit artikel stelt de auteur dat kinderen doorzettingsvermogen hebben omdat het een karaktereigenschap is. De meeste mensen hebben het vanuit zichzelf, een klein gedeelte van de mensen niet.
- ☐ 5. In dit artikel wordt gesteld dat de kinderen van nu geen doorzettingsvermogen meer hebben omdat hun ouders hen de hemel inprijzen en hen niet leren om met frustraties om te gaan.
- ☐ 6. In dit artikel stellen de auteurs dat kinderen wel doorzettingsvermogen hebben en dat de aard van het debat over de jeugd van tegenwoordig niet wezenlijk veranderd is. Elke generatie klaagt over de generatie die na haarzelf komt.
- ☐ 7. In dit artikel stellen de auteurs dat kinderen van nu geen doorzettingsvermogen hebben omdat alles voor hen geregeld wordt: ouders regelen met wie hun kinderen spelen, leerkrachten weten precies welke leerstof zij moeten aanbieden zodat het nét niet moeilijk genoeg is.
- ☐ 8. In dit artikel stellen de auteurs dat kinderen wel doorzettingsvermogen hebben en dat komt doordat de politiek hoge ambities en verwachtingen heeft m.b.t. het onderwijs. Kinderen leren vanzelf door te zetten door de hoge eisen in het onderwijs.

Een voorbeeld van een selectietaak met vier kaarten (Wason, 1968):

Letters en cijfers

Hieronder staan vier kaarten. Elke kaart heeft een letter aan de ene kant en een cijfer aan de andere kant. Jouw taak is om te bepalen welke kaarten je moet omdraaien om de volgende regel te toetsen:

"Als een kaart een klinker heeft aan de ene kant, dan staat er een even getal op de andere kant"

A	**D**	**4**	**7**

Je ziet alleen de letter OF het cijfer. Welke kaart of kaarten moet je omdraaien om te bepalen of de regel waar is of niet? Klik op de kaart of kaarten die jij wilt omdraaien. **Kies er bij dit soort opgaven telkens maximaal twee.**

In het eerste onderzoek (hoofdstuk 2) werd onderzocht of een instructie in een domeinspecifieke (onderwijs) context zou leiden tot betere prestaties op een confirmation bias-posttest in vergelijking met een instructie in een algemene context (dus onafhankelijk van een bepaald leerdomein zoals het basisonderwijs). 70 Pabostudenten werden willekeurig toegewezen aan een domeinspecifieke onderwijscontext (n = 37) of de algemene context (n = 33). Er werd verondersteld dat instructie en het oefenen van taken in een onderwijscontext de confirmation bias van pabostudenten meer zou verminderen dan instructie en oefenen in een algemene context vanwege een veronderstelde hogere interesse in de taak (Bangert-Drowns & Bankert, 1990; Heijltjes, Van Gog, Leppink, et al., 2014; Marin & Halpern, 2011; Niu et al., 2013). Alle deelnemers maakten een pretest met confirmation bias-taken. De deelnemers bestudeerden een instructievideo met voorbeelden in de domeinspecifieke onderwijscontext of een instructievideo met algemene voorbeelden (zie figuur 1 voor de QR-codes die leiden naar de instructievideo). Zij oefenden met confirmation bias-taken. De leermotivatie werd gemeten door middel van interesse in de taak. Daarom vulden alle deelnemers een vragenlijst in over hun interesse in de instructie en taken (topic interest; Mason et al., 2008; Schraw et al., 1995). De deelnemers maakten na de instructie een posttest, en vervolgens nogmaals een posttest na een week en na vijf weken, met gelijksoortige taken. De resultaten toonden aan dat de gemiddelde confirmation bias op de posttest na één week verminderde in vergelijking met de pretest, en dat pabostudenten die leerden en oefenden in een domeinspecifieke onderwijscontext een hogere interesse in de instructie en taken rapporteerden in vergelijking met degenen in de algemene context. Dit zou erop kunnen wijzen dat pabostudenten meer gemotiveerd zijn om te leren binnen een onderwijscontext dan leren in een algemene context.

Figuur 1. Links: Domeinspecifieke instructie (onderwijscontext). Rechts: Algemene instructie.

In het tweede onderzoek (hoofdstuk 3) is onderzocht of feedback in de vorm van uitgewerkte voorbeelden op oefentaken na een instructie over het verminderen van confirmation bias zou leiden tot betere prestaties op een confirmation bias-posttest en transfertest in vergelijking met feedback in de vorm van alleen het juiste antwoord op oefentaken, of geen feedback (Hattie, 2012; Hattie & Timperley, 2007; Shute, 2008; Vollmeyer & Rheinberg, 2005). Voor zover bekend, was dit de eerste studie die het effect van feedback op confirmation bias-taken onderzocht en de eerste studie om het differentiële effect van uitgewerkte feedback ten opzichte van feedback met het juiste antwoord te onderzoeken. Het onderzoek werd uitgevoerd onder 132 psychologie- en pabostudenten. Er werden vijf condities getest: (1) een instructievideo en oefentaken met feedback in de vorm van uitgewerkte voorbeelden (WE, "worked example"), (2) een instructievideo en oefentaken met feedback in de vorm van alleen het juiste antwoord (CA, "correct answer"), (3) een instructievideo en oefentaken zonder feedback (PO, "practice only"), (4) een instructievideo zonder oefening (VO, "video only") en (5) geen instructie en geen oefening (NT, "no treatment"). Condities 3, 4 en 5 fungeerden als controlecondities om respectievelijk de leereffecten van instructie en oefenen zonder feedback, alleen instructie en effecten van testen zonder interventie te bepalen. We veronderstelden dat WE > CA > PO > VO > NT de volgorde van leerprestatie en transfer van leren was. Alle deelnemers maakten eerst een confirmation bias-pretest en keken vervolgens naar de instructievideo (behalve de NT-conditie). Behalve de NT- en VO-conditie, oefenden de drie condities met confirmation bias-taken. In beide condities met feedback (WE en CA) kregen de deelnemers ofwel feedback in de vorm van uitgewerkte voorbeelden, ofwel feedback met alleen het juiste antwoord. In zowel de condities met feedback (WE en CA) als in de conditie zonder feedback (PO) rapporteerden de deelnemers vervolgens hun mentale inspanning. Ten slotte maakten alle deelnemers de post- en transfertest. De deelnemers aan beide feedbackcondities presteerden significant beter op isomorfe posttest taken in vergelijking met de condities zonder feedback. Er werden echter geen verschillen gevonden tussen de condities op transfer. Met uitzondering van de conditie zonder instructie en zonder oefening, werd in alle vier de condities een hogere score gevonden op de posttest vergeleken met de pretest. We vonden geen verschillen op de mentale inspanningsscores tussen de condities.

Het doel van de experimenten beschreven in hoofdstuk 4 en 5, was gericht op generatieve verwerkingsstrategieën. Er werd namelijk onderzocht of leren door lesgeven op video aan een fictief publiek (als vorm van zelfverklaren) gunstig was voor het leren van pabostudenten om hun confirmation bias te verminderen (hoofdstuk 4). In hoofdstuk 5 stond dezelfde vraag centraal maar was de uitkomstmaat het ontwerp van een lesplan voor burgerschapseducatie.

In hoofdstuk 4 wordt een experiment beschreven waarin het effect onderzocht is van verschillende generatieve verwerkingsstrategieën op de confirmation bias van pabostudenten (o.a., Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Kobayashi, 2019; Renkl, 1997; Roscoe & Chi, 2008). 141 Pabostudenten werden verdeeld over drie condities. In één conditie verwerkten de deelnemers de instructie door middel van het voorbereiden van een les met als doel de les vervolgens op video te geven aan een fictief publiek (TOV, teaching on video) en een conditie door alleen een les voor te bereiden (PTT, preparing to teach). In de controleconditie mochten de deelnemers zelf bepalen hoe ze de instructie die ze op papier kregen, wilden verwerken (CC, re-study controleconditie). Voor leren en transfer was de hypothese TOV > PTT > CC. Als metingen voor het leren werd de kwaliteit van de lesvoorbereidingen in kaart gebracht en samen met de confirmation bias- en transfertaken werd dat gebruikt om de effectiviteit van de verschillende verwerkingsstrategieën in kaart te brengen. Daarnaast werden vragenlijsten gebruikt en getest om de gevoelens van social presence en arousal te meten om eventuele verschillen tussen de generatieve verwerkingsstrategieën te verklaren. De resultaten toonden aan dat de gemiddelde confirmation bias in alle condities in dezelfde mate verminderde, en dat de kwaliteit van de lesvoorbereiding het hoogst was voor de TOV-deelnemers. De deelnemers die alleen een les voorbereidden (PTT) rapporteerden echter tegen de verwachtingen in gemiddeld hogere gevoelens van social presence in vergelijking met de lesgevende deelnemers. Alleen de lesgevende deelnemers rapporteerden gemiddeld hogere niveaus van arousal in vergelijking met de controleconditie. Er waren geen verschillen tussen de condities op transfer.

Het experiment dat is beschreven in hoofdstuk 5 was bedoeld om het experiment uit hoofdstuk 4 conceptueel te repliceren. De instructie werd gegeven aan 176 deelnemers in de TOV, PTT of CC conditie. In plaats van een instructie gericht op het verminderen van de confirmation bias was in dit experiment de instructie gericht op het onbevooroordeeld ("met zo min mogelijk confirmation bias") voorbereiden van een les binnen burgerschapseducatie. Hiervoor zijn bewustzijn en kennis over confirmation bias en perspectiefwisseling noodzakelijk. De kwaliteit van de uitleg van de instructie, alsmede gevoelens van social presence, arousal (alleen voor TOV en PTT), ruimdenkendheid en de conceptuele kennis van de instructie werden gemeten. Net als in het vorige onderzoek werd verwacht dat TOV > PTT > CC voor de kwaliteit van de lesvoorbereiding en de conceptuele kennis. We ontdekten dat de deelnemers gemiddeld hoger scoorden op ruimdenkendheid, maar deelnemers in de controleconditie maakten gemiddeld een significant nauwkeurigere en volledigere lesvoorbereiding dan de deelnemers die hun les gaven op video of de les voorbereidden. Een verkennende, algemene kwalitatieve analyse bracht geen verschillen aan het licht tussen de drie condities. Er werden geen verschillen gevonden tussen voorbereiden en aansluitend lesgeven en alleen voorbereiden op social presence en arousal. Dit resultaat was in overeenstemming met het experiment dat werd beschreven in hoofdstuk 4.

Conclusies

De resultaten van de vier experimenten dragen bij aan de theorievorming over het verminderen van de confirmation bias, met name hoe een instructie het beste kan worden ontworpen, en met welke instructie- en verwerkingsstrategieën er in het ontwerp rekening gehouden kan worden om het leren zo effectief mogelijk te laten verlopen. De bevindingen van het proefschrift benadrukken het belang van expliciete instructie om het leren te ondersteunen en de prestaties op verschillende soorten confirmation bias-taken te verbeteren (Adame, 2016; Beaulac & Kenyon, 2014; Correia, 2018; Dunbar et al., 2014; Dunbar et al., 2013; Heijltjes, 2014; Hirt & Markman, 1995; Korteling et al., 2021; Larrick, 2004; Lee et al., 2016; Lilienfeld et al., 2009; Lord et al., 1984; Morewedge et al., 2015; Mussweiler et al., 2000; Sellier et al., 2019; Van Peppen, 2020). De resultaten van het onderzoek ondersteunen het eerder gevonden positieve effect van feedback op leren in andere domeinen. In tegenstelling tot bevindingen van onder andere Butler et al. (2013) was de feedback met uitgewerkte voorbeelden in ons onderzoek even effectief als feedback met het juiste antwoord. Uit de derde studie (hoofdstuk 4) bleek dat de generatieve verwerkingsstrategie learning by teaching on video een positief effect heeft op de kwaliteit van de lesvoorbereiding, maar niet op de confirmation bias-taken. In het vierde experiment (hoofdstuk 5) werd geen leereffect van generatieve verwerking gevonden op de mate van ruimdenkendheid van een lesvoorbereiding.

De studies in dit proefschrift vonden - in tegenstelling tot andere bevindingen in de literatuur - geen bewijs dat domeinspecificiteit, uitgebreide feedback en leren door les te geven op video resulteerden in een extra reductie van de confirmation bias ten opzichte van een controleconditie (Abrami et al., 2015; Bangert-Drowns & Bankert, 1990; Bangert-Drowns et al., 1991; Butler et al., 2013; Ennis, 1989; Fiorella & Mayer, 2013, 2014; Heijltjes, Van Gog, & Paas, 2014; Hoogerheide et al., 2016; Hoogerheide, Renkl, et al., 2019; Marin & Halpern, 2011; Niu et al., 2013; Shute, 2008). Hieronder worden deze verrassende bevindingen één voor één besproken.

Domeinspecificiteit. In de eerste studie, beschreven in hoofdstuk 2, was de interventie kort, wat mogelijk een positief effect van domeinspecificiteit op prestaties heeft belemmerd. (Lorencová et al., 2019). Mogelijk zou er wel een prestatieverschil zichtbaar kunnen worden bij een langere instructiereeks. Het verschil in volgehouden aandacht en tijd voor de taak tussen een domeinspecifieke instructie en een algemene instructie zal waarschijnlijk toenemen ten gunste van de domeinspecifieke conditie, wat dan kan resulteren in een prestatievoordeel. Het is echter mogelijk dat een door interesse gedreven aanhoudende aandacht en extra tijd voor de taak niet voldoende zijn om de prestaties te verbeteren. Wellicht is het mogelijk om de instructie en oefening aan te vullen met feedback, bedoeld om studenten te helpen hun conceptueel begrip van de leerstof te verbeteren. Voor toekomstige onderzoek is het interessant om deze verwachte causaliteit specifiek te bekijken voor de confirmation bias.

Feedback. Op basis van de feedbackliteratuur (Hattie & Timperley, 2007; Roelle et al., 2017; Van der Kleij et al., 2015; Vollmeyer & Rheinberg, 2005), kunnen er verschillende redenen zijn waarom uitgebreide feedback in het tweede experiment, beschreven in hoofdstuk 3 van dit proefschrift, niet effectiever was dan feedback met alleen het correcte antwoord. Een mogelijkheid is dat deelnemers de uitgebreide feedback in de uitgewerkte voorbeeldconditie verwerkten, de uitwerking begrepen, maar deze misschien niet nodig hadden om hun prestaties te verbeteren. In dit geval is feedback niet effectief en zullen de deelnemers meestal de feedbackboodschap negeren (Hattie & Timperley, 2007; Roelle et al., 2017; Van der Kleij et al., 2015). Misschien maakte het juiste antwoord (dat in beide condities werd gegeven) het voor deelnemers mogelijk om een heuristiek af te leiden, wat voldoende was om tot betere prestaties te komen voor het soort taken dat we hen aanboden. Deelnemers kunnen bijvoorbeeld uit het juiste antwoord op de selectietaak met vier kaarten hebben afgeleid dat ze de kaart die de regel bevestigt en de kaart die de regel ongeldig maakt, moeten omdraaien. Het gebruik van deze vuistregel zal de prestaties verbeteren, waardoor het potentiële voordeel van extra begrip, dat mogelijk het gevolg is van het verwerken van de uitgebreide feedback, wordt beperkt (of zelfs teniet wordt gedaan). Een andere mogelijkheid is dat deelnemers de uitgebreide feedback in de uitgewerkte voorbeeldconditie hebben verwerkt, maar de gegeven uitleg niet begrepen. Hierdoor hadden ze misschien aan de hand van het juiste antwoord - mogelijk - een vuistregel bedacht voor het oplossen van dit soort taken. In deze situatie zou het gebrek aan begrip van de feedback onze resultaten verklaren. Ten slotte is het mogelijk dat deelnemers de uitgebreide feedback in het uitgewerkte voorbeeld niet hebben verwerkt. Dat kan te wijten zijn aan verschillende redenen die verband houden met motivatie of een gebrek aan de noodzaak die deelnemers al dan niet voelden om het uitgewerkte voorbeeld door te nemen. In elk geval zal dit tot een gebrek aan verschil tussen de twee feedbackcondities leiden. Deze laatste verklaring wordt enigszins ondersteund door onze gegevens want in het onderzoek verschilden de twee condities niet in tijd die werd besteed aan de taak. Dit kan betekenen dat de uitgebreide feedback oppervlakkig is verwerkt. Het zou er echter ook op kunnen duiden dat deelnemers in de conditie met alleen het correcte antwoord nadachten over de rationale achter het juiste antwoord en dat dit het effect van de uitgewerkte voorbeeldconditie verminderde. Kortom, op dit moment is het alleen mogelijk om te speculeren over waarom de twee feedbackcondities niet verschilden in prestaties. Om het (de) onderliggende mechanisme(n) achter het gebrek aan effect te achterhalen, moet er meer inzicht komen in hoe studenten de feedback zowel kwalitatief als kwantitatief verwerken.

Leren door lesgeven op video. In de twee studies over generatieve verwerking die worden beschreven in hoofdstuk 4 en 5 werd geen bewijs gevonden voor de social presence hypothese. De conditie met de veronderstelde hoogste social presence, namelijk lesgeven op video (TOV), verschilde niet op gemiddelde arousal in vergelijking met alleen het voorbereiden om les te geven (PTT). Deelnemers aan beide condities moesten hun les echter op papier voorbereiden. Na afloop moesten de TOV-deelnemers hun uitleg voorlezen op de camera van hun eigen smartphone of laptop. Dit zou kunnen hebben geresulteerd in kennisreproductie in plaats van de beoogde kennisopbouw, wat weinig extra impact heeft gehad op de uiteindelijke leerprestaties. Een andere verklaring voor het ontbreken van prestatieverschillen zou de kwaliteit van de zelfverklaringen kunnen zijn. Deze kwaliteit was in beide experimenten niet bijzonder hoog in beide condities (TOV en PTT). Effecten van zelfverklaring op leren en prestaties zijn echter afhankelijk van een voldoende hoge kwaliteit van zelfverklaringen. Het is mogelijk dat het begripsniveau van de deelnemers na instructie en oefening niet hoog genoeg was om een gunstig effect te hebben van lesgeven op video en het voorbereiden om les te geven (zie bijvoorbeeld Jacob et al., 2021).

Dat er geen verschil is gevonden tussen de condities op social presence kan wellicht verklaard worden door de manier waarop we hebben gemeten, namelijk door zelfrapportage na de interventie. Gevoelens van social presence zouden kunnen zijn afgenomen omdat het eigenlijke lesgeven al voorbij was. Het meten van social presence tijdens de leeractiviteit kan uitkomst bieden. Een voorbeeld is het gebruik van polsbandjes die een verandering in de elektrische weerstand of temperatuur van de huid meten die wordt veroorzaakt door spanning (Biocca & Harms, 2002, October; Cui, 2013; Gunawardena & Zittle, 1997; Hoogerheide, Renkl, et al., 2019). Een ander voorbeeld is het tellen van het aantal persoonlijke referenties in de toelichting (Hoogerheide et al., 2016; Jacob et al., 2020; Lachner et al., 2018). Het ontbreken van een algemene overeenstemming over hoe social presence moet worden gemeten, beperkt de vergelijkbaarheid tussen studies en daaropvolgende verklaringen over waarom het wel of niet bijdraagt aan leren. Al met al is het nog steeds niet helemaal duidelijk wanneer verhoogde social presence al dan niet wenselijk is voor leren (Lachner et al., 2021), vooral in de context van generatieve verwerking bij beginners (Jacob et al., 2021). Verder onderzoek is nodig om te bepalen wanneer en waarom gevoelens van social presence gunstig zijn voor leren, en in welke leercontexten.

Bijdrage aan Open Science

Een sterk punt van dit proefschrift is dat drie studies vooraf waren geregistreerd op het Open Science Framework (OSF). Pre-registratie op het OSF hield voor de studies uit hoofdstuk 3, 4 en 5 van dit proefschrift in dat de hypothesen, de gebruikte materialen, de geplande dataverzamelingen, de geplande methoden van onderzoek en de geplande data-analyses van de onderzoeken vrij toegankelijk en beschikbaar zijn voor iedereen. Door pre-registratie nemen onderzoekers in een vroeg stadium belangrijke beslissingen en deze beslissingen zijn transparant (Nosek et al., 2018). Pre-registratie is belangrijk om enkele twijfelachtige maar soms onbedoelde onderzoekspraktijken zoals HARKing (Hypothesizing After the Results are Known; Kerr, 1998; Simmons et al., 2011) of data dredging te verminderen. Deze laatste term is ook bekend als *p*-hacking, het grasduinen in opgehaalde onderzoeksdata op zoek naar patronen die statistisch significant zijn (Munafò et al., 2017; Nosek & Lakens, 2014). Pre-registratie zal vals-positieven in onderzoek verminderen, omdat dit de vrijheidsgraden van onderzoekers beperkt (Nosek et al., 2018; Wicherts et al., 2016). Pre-registratie kan helpen om "de replicatiescrises om te zetten in een geloofwaardigheidsrevolutie" (vertaald vanuit Plucker & Makel, 2021, p. 91), vertrouwen in wetenschappelijke bevindingen te versterken en onderwijspsychologisch onderzoek sneller en efficiënter te ontwikkelen.

Aanbevelingen voor de onderwijspraktijk

De studies zijn gedaan in een gecontroleerde setting met een specifiek soort taken die niet altijd representatief zijn voor het werk dat leraren in de praktijk uitvoeren. Desalniettemin bieden de resultaten aanknopingspunten voor de onderwijspraktijk. De resultaten onderstrepen het belang van instructie in combinatie met oefening en met feedback. Daarnaast laten hoofdstuk 4 en 5 verrassende resultaten zien met betrekking tot het leren van lesgeven op video. Deze strategie lijkt veelbelovend te zijn voor de onderwijspraktijk omdat de leeractiviteit nauw aansluit bij wat leraren normaal gesproken doen in de klas. Het is voor leraren belangrijk om zich te realiseren dat deze de strategie niet zonder meer effectief is, zoals ook uit ons onderzoek blijkt. Meer onderzoek is nodig om te weten te komen wat wanneer wel of niet werkt bij het leren verminderen van de confirmation bias. Een ander inzicht dat voortvloeit uit ons onderzoek en van belang is voor de praktijk, is dat uitgebreide feedback niet altijd beter werkt dan alleen het juiste antwoord als feedback geven. Hiermee kan ook tijdens het lesgeven rekening gehouden worden.

Naast de aanbevelingen die direct volgen uit de onderzoeken van dit proefschrift, zijn er in de literatuur, die deels ten grondslag ligt aan de studies in dit experiment, ook algemene aanbevelingen voor het onderwijs in kritisch denken. Zo stellen Abrami et al. (2015) voor dat de instructie een combinatie van het voeren van een dialoog, het aanbieden van mentorschap en het geven van een authentieke instructie moet bevatten omdat deze drie factoren samen als een katalysator werken bij kritisch leren denken. Methoden hiervoor zijn bijvoorbeeld klassikale debatten en discussies, socratische dialogen, rollenspellen, casestudies, simulaties, games en natuurlijk de stages. Good practices van studies in de onderwijspraktijk over kritisch denken zijn te vinden in studies door bijv. Bell en Loon (2015), Dumitru (2012), Hill (2000), Tiruneh et al. (2016), Van Peppen (2020) en Zohar en Schwartzer (2005). Voor een succesvolle implementatie in de praktijk is het van groot belang dat de leraren zichzelf blijven professionaliseren, dat ze positief staan tegenover de relevantie van het verminderen van de confirmation bias door middel van perspectiefwisseling en dat zij het besef hebben dat kritisch denken geen bijvangst is van het onderwijs, maar iets dat tijd en aandacht vraagt (Janssen, Mainhard, et al., 2019; Janssen, Meulendijks, et al., 2019). Het delen van informatie en good practices kan hierin een belangrijke rol spelen (bijvoorbeeld via het online platform Kritisch Leren Denken: kritischdenkenhbo.nl).

De meeste deelnemers aan de experimenten in dit proefschrift waren pabostudenten. Zij zullen in hun toekomstige baan als leraar op de basisschool zelf kritisch moeten leren denken maar dit ook onderwijzen aan de leerlingen in hun klas. Dit is belangrijk omdat kritisch denken bijdraagt aan onder andere de persoonlijke ontwikkeling en aan goed burgerschap. Op basis van de resultaten, kan worden aangenomen dat de leraar een cruciale rol speelt omdat instructie en feedback voorwaardelijk zijn om kritisch te leren denken, ook op de basisschool. Daarom is het voorstel om het leren en het leren onderwijzen van kritisch denken te verweven in het pabocurriculum en om verder te onderzoeken hoe jonge kinderen effectief kritische denkvaardigheden kunnen leren. Om leraren in de onderwijspraktijk te ondersteunen bij het ontwerpen van curricula waarin kritisch denkvaardigheden expliciet worden behandeld, worden in elk geval cursussen voor professionele ontwikkeling (in dit geval voor lerarenopleiders) aanbevolen (Janssen, Mainhard, et al., 2019).

Er zijn veelbelovende ontwerpprincipes voor een kritisch denken instructie, maar er is meer onderzoek nodig om meer te weten te komen onder welke omstandigheden deze het meest effectief zijn voor leren. Instructie in een domeinspecifieke leercontext, of het gebruik van de generatieve leerstrategie teaching on video hebben de potentie om het leren te verbeteren, maar er is aanvullend onderzoek nodig om beter inzicht te krijgen in de leermechanismen die aan deze strategieën ten grondslag liggen. Ook hebben leraren de verantwoordelijkheid om hun leerlingen in de klas tot goede kritische denkers te scholen. Daarom moeten leraren het belang inzien van het bevorderen van kritisch denken bij jonge kinderen en vertrouwen hebben in hun eigen kritische denkcapaciteiten (Lorencová et al., 2019). Kortom, elke leraar moet kritisch kunnen denken, hiervoor voldoende toegerust zijn en zich competent genoeg voelen om kritische denkvaardigheden te onderwijzen aan de leerlingen.

Curriculum Vitae

Suzan van Brussel (1975), obtained her Master's degree in Linguistics and Cultural Studies at Tilburg University in 1999. In 2001, she graduated from Fontys University of Applied Sciences as a primary teacher. She started her career as a teacher and deputy director at various primary schools. Since 2011, she has been working as a teacher educator at the primary teacher institute of Avans University of Applied Sciences in Breda. In 2013, she was voted Avans Lecturer of the Year. Suzan held various positions within the primary teacher institute, such as language didactics lecturer, research lecturer, chair of the assessment committee and member of the curriculum committee. In August 2015, she decided to combine working as a teacher educator with a part time PhD trajectory at the Department of Psychology, Education and Child Studies at the Erasmus University Rotterdam, resulting in this dissertation. During her PhD trajectory, Suzan was member of the Leerkracht research group of the primary teacher institute and she presented her work at various (inter)national conferences. She gave workshops and presentations to teacher educators, student teachers and primary school educational professionals in which she translated her findings from this research project to the educational practice. She designed several (online) courses aimed at learning to think critically for student teachers and primary school teachers. From 2021 onwards, next to being a teacher educator, Suzan works as a senior researcher in the Digital Didactics research group at Avans University of Applied Sciences with a research focus on the optimal guidance of students' decision-making processes in a flexible higher education curriculum. Suzan is highly committed to the implementation of the Dutch Code of Conduct for Scientific Integrity in her role as the confidential counsellor for scientific integrity for researchers, teachers and students of Avans University of Applied Sciences. In 2022 she started her own company "Suzan Et Aliae" from which she collaborates on research projects as a freelancer. In addition to her work, she is a member of the Supervisory Board of the PrentenboekenPlus Foundation. This foundation supports educational professionals with picture books that appeal to different senses and are therefore particularly suitable for children with visual, auditory and communicative disabilities.

Publications, presentations, and workshops

2022 Van Brussel, S. (2022, January). De andere kant van het gelijk. Studio Wetenschap.

2021 Van Brussel, S. (2021, August). Comparing instructional strategies to support student teachers' learning to prepare an open-minded citizenship lesson. In L. Jacob & A. Lachner, Effectiveness of Teaching: Does Task-Design matter? Symposium conducted at the biannual conference of the European Association for Research on Learning and Instruction (EARLI), online through the University of Göteborg, Sweden.

Van Brussel, S. (2021, September). Een open-minded leerkracht, open-minded leerlingen? Workshop conducted at the completion of the Research Group Leer*kracht*, Avans University of Applied Sciences, Breda, the Netherlands.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2021). Teaching on video as an instructional strategy to reduce confirmation bias - A pre-registered study. *Instructional Science, (20210626)* https://doi.org/10.1007/s11251-021-09547-4.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (revised manuscript submitted). "Consider the Opposite" –The Effects of Learning Context on Student Teachers' Confirmation Bias and Topic Interest.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (revised manuscript submitted). Comparing instructional strategies to support student teachers' learning to prepare an open-minded citizenship education lesson.

Van Brussel, S. (2021). Cursus Kritisch Denken, recorded for the E-wise online courses platform for primary school teachers. Available after subscription from e-wise.nl.

2020 Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2020). "Consider the Opposite"– Effects of elaborative feedback and correct answer feedback on reducing confirmation bias – A pre-registered study. *Contemporary Educational Psychology*, 60, 101844. https://doi.org/10.1016/j.cedpsych.2020.101844.

Van Brussel, S. (2020, August). Teaching on Video as an Instructional Strategy to Reduce Confirmation Bias. In V. Hoogerheide & J. Roelle, When and why does learning by teaching (not) foster learning? Symposium conducted at the biannual conference of Special Interest Groups 6 and 7 (Instructional Design & Technology Enhanced Learning and Instruction) of the European Association for Research on Learning and Instruction (EARLI SIG 6/7 2020), online through the University of Tubingen, Germany.

2019 Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2019, April). Effects of elaborative feedback and correct answer feedback on reducing confirmation bias. A pre-registered study. Poster session, presented at the biannual American Educational Research Association (AERA), Toronto, Canada.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2019, March). Effects of elaborative feedback and correct answer feedback on reducing confirmation bias. A pre-registered study. Paper session, presented at the DPECS Graduate Research Day, Erasmus University Rotterdam, The Netherlands.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2019, March). Kritisch denken vraagt om perspectiefwisseling – een onderzoek op de pabo van Avans Hogeschool (Critical thinking requires perspective taking – a study at the primary teacher institute of Avans University of Applied Sciences). Paper session, presented at the annual meeting of Velon Velov, meeting for teacher educators, Breda, The Netherlands.

2018 Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2018, June). Ondersteboven denken: het effect van de leercontext en feedback op het leren toepassen van perspectiefwisseling (Thinking upside down: the effect of learning context and feedback on learning to apply perspective shifting). Paper session, presented at the biannual meeting of the Educational Research Days (ORD), Nijmegen, The Netherlands.

2017 Van Brussel, S., Timmermans, M. C. L., Verkoeijen P. P. J. L., & Paas, F. (2017, September). The effect of domain-specificity in a video-based instruction on reducing student teachers' confirmation bias. Poster session, presented at the biannual meeting of the European Association for Research on Learning and Instruction (EARLI), Tampere, Finland. DOI: 10.13140/RG.2.2.31178.88000.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2017, November). The effect of domain-specificity in a video-based instruction on reducing student teachers' confirmation bias. Pitch at the DPECS Graduate Research Day, Drawing attention, Erasmus University Rotterdam, the Netherlands.

2016 Van Brussel, S. (2016, January). De leerkracht van kritisch denken – de rol van de leerkracht (The learning power of critical thinking, the role of the teacher). Workshop conducted at the kickoff of the Research Group Leerkracht, Avans University of Applied Sciences, Breda, the Netherlands.

Van Brussel, S. (2016, February). De leerkracht van kritisch denken – de rol van de leerkracht (The learning power of critical thinking, the role of the teacher). Workshop for primary school teachers at Avans University of Applied Sciences, Breda, The Netherlands.

Van Brussel, S. (2016, March). Bezieling is…blijven nadenken (Inspiration is…keep thinking.). Workshop at the symposium Wat bezielt de leerkracht? (What inspires the teacher?) for teacher educators and primary school teachers at Avans University of Applied Sciences, Breda, The Netherlands.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2016, May). Effecten van “video based modeling examples” op de kritische denkvaardigheden van pabostudenten (Effects of video based modeling examples on student teachers’ critical thinking skills). Poster session, presented at the biannual meeting of the Educational Research Days (ORD), Rotterdam, The Netherlands.

2015

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2015, November). Teaching and learning critical thinking skills: effective instructional designs to enhance critical thinking of trainee teachers and pupils in primary education. Round table session at biannual meeting of the European Association for Practitioner Research on Improving Learning (EAPRIL), Bel-Val, Luxembourg.

References

Ab Kadir, M. A. (2017). What teacher knowledge matters in effectively developing critical thinkers in the 21 st century curriculum? *Thinking Skills and Creativity*, *23*, 79-90. https://doi.org/10.1016/j.tsc.2016.10.011

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, A. C., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, *85*(2), 275-314. https://doi.org/10.3102/0034654314551063

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*(4), 1102-1134. https://doi.org/10.3102/0034654308326084

Adame, B. J. (2016). Training in the mitigation of anchoring bias: A test of the consider-the-opposite strategy. *Learning and Motivation*, *53*, 36-48.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*(3), 545-561.

Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications*, *2*(1), 39-47.

Bacon, F., Jardine, L., & Silverthorne, M. (1620 / 2000). *Francis Bacon: the new organon*. Cambridge University Press. (1620)

Baehr, J. (2011). The structure of open-mindedness. *Canadian Journal of Philosophy*, *41*(2), 191-213.

Bangert-Drowns, R. L., & Bankert, E. (1990, April 16-20, 1990). *Meta-analysis of effects of explicit instruction for critical thinking* American Educational Research Association, Boston, MA.

Bangert-Drowns, R. L., Kulik, C., Kulik, J., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213-238.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637. https://doi.org/10.1037//0033-2909.128.4.612

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking & Reasoning*, *1*(3), 221-235. https://doi.org/10.1080/13546789508256909

Baron, J. (2008). Actively open-minded thinking. In J. Baron (Ed.), *Thinking and Deciding* (4th ed., pp. 199-232). Cambridge University Press.

Bartlett, F. C. (1932). *Remembering: a study in experimental and social psychology*. Cambridge University Press.

Beaulac, G., & Kenyon, T. (2014). Critical thinking education and debiasing. *Informal Logic*, *34*(4), 341-363.

Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching and Learning*, *8*(2), 25-42.

Bell, R., & Loon, M. (2015). The impact of critical thinking disposition on learning using business simulations. *International Journal of Management Education*, *13*(2), 119-127. https://doi.org/10.1016/j.ijme.2015.01.002

Berthet, V. (2021). The measurement of individual differences in cognitive biases: A review and improvement. *Frontiers in Psychology*, *12*, 630177. https://doi.org/10.3389/fpsyg.2021.630177

Biocca, F., & Harms, C. (2002, October). *Defining and measuring social presence: Contribution to the networked minds theory and measure.* 5th Annual International Workshop on Presence, Porto, Portugal.

Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: comparing intended versus actual study behaviour. *Memory*, *25*(6), 784-792. https://doi.org/10.1080/09658211.2016.1221974

Bown, A. (2017). Elaborative feedback to enhance online second language reading comprehension. *English Language Teaching*, *10*(12), 164-171.

Boyle, G. J., Saklofske, D. H., & Matthews, G. (2015). *Measures of personality and social psychological constructs*. Academic Press.

Brandsford, J. D., Pellegrino, J. W., & Donovan, S. (1999). *How people learn: bridging research and practice*. National Academy Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42.

Bruine de Bruin, W., Parker, A., & Fischhoff, B. (2007). Individual differences in adult decision making competence. *Journal of Personality and Social Psychology*, *92*(5), 938-956. https://doi.org/10.1037/0022-3514.92.5.938

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290-298.

Candel, C., Máñez, I., Cerdán, R., & Vidal-Abarca, E. (2021). Delaying elaborated feedback within computer-based learning environments: The role of summative and question-based feedback. *Journal of Computer Assisted Learning*, *37*(4), 1015-1029. https://doi.org/10.1111/jcal.12540

Catapano, R., Tormala, Z. L., & Rucker, D. D. (2019). Perspective taking and self-persuasion: Why "putting yourself in their shoes" reduces openness to attitude change. *Psychological Science*, *30*(2), 1-12. https://doi.org/10.1177/0956797618822697

Cavojova, V., Srol, J., & Adamus, M. (2018). My point is valid, yours is not: myside bias in reasoning about abortion. *Journal of Cognitive Psychology*, *30*(7), 656-669. https://doi.org/10.1080/20445911.2018.1518961

Chambers, C., & Munaf, M. (2013). Trust in science would be improved by study pre-registration. https://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 151-155.

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, *104*, 1-23. https://doi.org/10.1016/j.geb.2017.02.015

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73-105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145-182. https://doi.org/10.1207/s15516709cog1302_1

Chi, M. T. H., Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477. https://doi.org/10.1207/s15516709cog1803_3

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *Journal of the Learning Sciences*, *6*(4), 347-365. https://doi.org/10.1207/s15327809jls0604_1

Conlin, L. D., Kuo, E., & Hallinen, N. R. (2019). How null results can be significant for physics education research. *Physical Review Physics Education Research*, *15*(2), 02010401 - 02010413. https://doi.org/https://doi.org/10.1103/PhysRevPhysEducRes.15.020104

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*(4), 347-362.

Correia, V. (2018). Contextual debiasing and critical thinking: Reasons for optimism. *Topoi: An International Review of Philosophy*, *37*(1), 103-111. https://doi.org/10.1007/s11245-016-9388-x

Cui, G. (2013). Evaluating online social presence: An overview of social presence assessment. *Journal of Educational Technology Development and Exchange*, *6*(1), Article 3. https://doi.org/10.18785/jetde.0601.02

Danielson, R. W., & Sinatra, G. M. (2017). A relational reasoning approach to text-graphic processing. *Educational Psychology Review*, *29*(1), 55-72.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20-33.

De Boer, H., Bosker, R. J., & van der Werf, M. P. C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, *102*(1), 168-179.

De Boer, H., Timmermans, A. C., & van der Werf, M. P. C. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: narrative review and meta-analysis. *Educational Research and Evaluation*, *24*(3-5), 180-200. https://doi.org/10.1080/13803611.2018.1550834

De Neys, W. (2021). On dual- and single-process models of thinking. *Perspectives on Psychological Science*, 1-16. https://doi.org/10.1177/1745691620964172

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Dirkx, K. J. H., Camp, G., Kester, L., & Kirschner, P. A. (2019). Do secondary school students make use of effective study strategies when they study on their own? *Applied Cognitive Psychology*, *33*(5), 952-957. https://doi.org/10.1002/acp.3584

Douglas, N. L. (2000). Enemies of critical thinking: Lessons from social psychology research. *Reading Psychology*, *21*(2), 129-144.

Dumitru, D. (2012). Critical thinking and integrated programs. The problem of transferability. *Procedia - Social and Behavioral Sciences*, *33*, 143-147. https://doi.org/10.1016/j.sbspro.2012.01.100

Dunbar, N. E., Miller, C. H., Adame, B. J., Elizondo, J., Wilson, S. N., Lane, B. L., . . . Zhang, J. (2014). Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game. *Computers in Human Behavior*, *37*, 307-318. https://doi.org/10.1016/j.chb.2014.04.053

Dunbar, N. E., Wilson, S. N., Adame, B. J., Elizondo, J., Jensen, M. L., Miller, C. H., . . . Squire, K. (2013). MACBETH: Development of a training game for the mitigation of cognitive bias. *International Journal of Game-Based Learning*, *3*(4), 7-26. https://doi.org/10.4018/ijgbl.2013100102

Duncan, T., & Pintrich, P. (1992). *Critical thinking and its relationship to motivation, learning strategies, and classroom experience* AERA, Washington, DC.

Dunlosky, J., Rawson, K. A., Marsh, E. J., 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*(1), 4-58. https://doi.org/10.1177/1529100612453266

Duran, D. (2017). Learning-by-teaching. Evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International*, *54*(5), 476-484. https://doi.org/10.1080/14703297.2016.1156011

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, *18*(3), 4-10.

Evans, J. S. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, *87*(2), 223-240. https://doi.org/10.1111/j.2044-8295.1996.tb02587.x

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454-459.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: the psychology of deduction*. Lawrence Erlbaum.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Facione, P. A. (1990). *Executive summary of critical thinking: A statement of expert consensus for purposes of educational assessment and instruction, including all tables, findings and recommendations of The Delphi Report*. California Academic Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Finn, B., Thomas, R., & Rawson, K. A. (2018). Learning more from feedback: Elaborating feedback with examples enhances concept learning. *Learning and Instruction*, *54*, 104-113.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(4), 281-288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, *39*(2), 75-85. https://doi.org/10.1016/j.cedpsych.2014.01.001

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717-741.

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, *89*, 465-470. https://doi.org/https://doi.org/10.1016/j.chb.2018.07.015

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology*, *112*(5), 895-906. https://doi.org/10.1037/edu0000408

Fonseca, B. A., & Chi, M. T. H. (2013). Instruction based on self-explanation. In *Handbook of Research on Learning and Instruction* (pp. 310-335). Routledge. https://doi.org/10.4324/9780203839089.ch15

Foster, G. C., Schmidt, C. R., & Sabatino, D. (1976). Teacher expectancies and the label "Learning disabilities". *Journal of Learning Disabilities*, *9*(2), 111-114. https://doi.org/10.1177/002221947600900209

Galinsky, A. D., Maddux, W. W., Gilin, D., & White, J. B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science*, *19*(4), 378-384. https://doi.org/10.1111/j.1467-9280.2008.02096.x

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J., & Tremblay, R. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development*, *87*(1), 165-175. https://doi.org/10.1111/cdev.12458

Giesbers, S. J. H., Rienties, B. C., Tempelaar, D. T., & Gijselaers, W. H. (2014). Why increased social presence through web videoconferencing does not automatically lead to improved learning. *E-Learning and Digital Media*, *11*(1), 31-45.

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

Grossnickle, E. M., Dumas, D., Alexander, P. A., & Baggetta, P. (2016). Individual differences in the process of relational reasoning. *Learning and Instruction*, *42*, 141-159. https://doi.org/10.1016/j.learninstruc.2016.01.013

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, *1*(2), 147-166.

Gunawardena, C. N., & Zittle, F. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, *11*(3), 8-26.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449-455.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: are self-testing and scheduling related to achievement? *Psychonomic Bulletin and Review*, *19*(1), 126–134.

Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 724-746). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470939376.ch25

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hattie, J. (2015). The applicability of Visible learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79-91.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99-136.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112.

Heijltjes, A. (2014). *Cultivating critical thinking: The effects of instructions on economics students' reasoning* Erasmus University Rotterdam]. WorldCat. Rotterdam.

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. https://doi.org/10.1016/j.learninstruc.2013.07.003

Heijltjes, A., Van Gog, T., & Paas, F. (2014). Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology*, *28*(4), 518-530.

Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review*, *13*(3), 191-209.

Hill, L. (2000). What does it take to change minds? Intellectual development of preservice teachers. *Journal of Teacher Education*, *51*(1), 50-62.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, *69*(6), 1069-1086.

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heijltjes, A., & Van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology*, *44-45*, 95-106. https://doi.org/10.1016/j.cedpsych.2016.02.005

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014a). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, *41*, 80-91.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014b). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction*, *33*, 108-119.

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & Van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, *111*(1), 45-56. https://doi.org/10.1037/edu0000272

Hoogerheide, V., Visee, J., Lachner, A., & Van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction*, *64*, 1-13. https://doi.org/10.1016/j.learninstruc.2019.101226

Hornstra, T. E., Stroet, K., van Eijden, E., Goudsblom, J., & Roskamp, C. (2018). Teacher expectation effects on need-supportive teaching, student motivation, and engagement: A self-determination perspective. *Educational Research and Evaluation*, *24*(3-5), 324-345.

Huang, G. C., Lindell, D., Jaffe, L. E., & Sullivan, A. M. (2016). A multi-site study of strategies to teach critical thinking: 'Why do you think that?'. *Medical Education*, *50*(2), 236-249. https://doi.org/10.1111/medu.12937

Jacob, L., Lachner, A., & Scheiter, K. (2020). Learning by explaining orally or in written form? Text complexity matters. *Learning and Instruction*, *68*, 101344. https://doi.org/10.1016/j.learninstruc.2020.101344

Jacob, L., Lachner, A., & Scheiter, K. (2021). Does increasing social presence enhance the effectiveness of writing explanations? *PloS one*, *16*(4), e0250406. https://doi.org/10.1371/journal.pone.0250406

Janssen, E. M., Mainhard, T., Buisman, R. S. M., Verkoeijen, P. P. J. L., Heijltjes, A., Van Peppen, L. M., & Van Gog, T. (2019). Training higher education teachers' critical thinking and attitudes towards teaching it. *Contemporary Educational Psychology*, *58*, 310–322. https://doi.org/https://doi.org/10.1016/j.cedpsych.2019.03.007

Janssen, E. M., Meulendijks, W., Mainhard, T., Verkoeijen, P. P. J. L., Heijltjes, A. E. G., Van Peppen, L. M., & Van Gog, T. A. J. M. (2019). Identifying characteristics associated with higher education teachers' Cognitive Reflection Test performance and their attitudes towards teaching critical thinking. *Teaching and Teacher Education vol. 84, pp. 139-149*. http://repub.eur.nl/pub/124818

Janssen, E. M., Verkoeijen, P. P. J. L., Heijltjes, A., Mainhard, T., Van Peppen, L. M., & Van Gog, T. (2020). Psychometric properties of the Actively Open-minded Thinking scale. *Thinking Skills and Creativity*, *36*, 100659. https://doi.org/10.1016/j.tsc.2020.100659

Johnson, R. H., & Hamby, B. (2015). A meta-level approach to the problem of defining 'critical thinking'. *Argumentation*, *29*(4), 417-430. https://doi.org/10.1007/s10503-015-9356-4

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557-571.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*(4), 237-251.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111-117. https://doi.org/10.1177/001316447403400115

Karpicke, J., Butler, A., & Roediger III, H. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471-479.

Kellen, D., & Klauer, K. C. (2019). Theories of the Wason selection task: A critical assessment of boundaries and benchmarks. *Computational Brain & Behavior*, *3*(3), 341-353. https://doi.org/10.1007/s42113-019-00034-1

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, *18*(4), 431-439.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196-217.

Kobayashi, K. (2019). Learning by preparing-to-teach and teaching: A meta-analysis. *Japanese Psychological Research*, *61*(3), 192-203. https://doi.org/10.1111/jpr.12221

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory*, *6*(2), 107-118. https://doi.org/10.1037//0278-7393.6.2.107

Korteling, J. E., Gerritsma, J., & Toet, A. (2021). *Retention and transfer of cognitive bias mitigation interventions: A systematic literature study*. TNO Human Factors.

Kwong, J. M. C. (2016). Open-mindedness as engagement. *The Southern Journal of Philosophy*, *54*(1), 70-86. https://doi.org/10.1111/sjp.12163

Lachner, A., Backfisch, I., Hoogerheide, V., Van Gog, T., & Renkl, A. (2019). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology*, *112*(4), 1-13. https://doi.org/10.1037/edu0000396

Lachner, A., Hoogerheide, V., van Gog, T., & Renkl, A. (2021). Learning-by-teaching without audience presence or interaction: When and why does it work? *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09643-4

Lachner, A., Ly, K.-T., & Nückles, M. (2018). Providing written or oral explanations? Differential effects of the modality of explaining on students' conceptual learning and transfer. *Journal of Experimental Education*, *86*(3), 344-361.

Lai, E. R. (2011). *Critical thinking: A literature review* (1 ed., Vol. 6). Pearson Education.

Lakens, D., Haans, A., & Koole, S. L. (2012). Één onderzoek is géén onderzoek: het belang van replicaties voor de psychologische wetenschap. *De Psycholoog: Maandblad van het Nederlands Instituut van Psychologen*, *47*(9), 10-18.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Larrick, R. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 316-338). Blackwell Publishing Ltd.

Lee, Y. H., Dunbar, N. E., Miller, C. H., Lane, B. L., Jensen, M. L., Bessarabove, E., . . . Wilson, S. N. (2016). Training anchoring and representativeness bias mitigation through a digital game. *Simulation & Gaming*, *47*(6), 751-779. https://doi.org/10.1177/1046878116662955

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*(4), 390-398.

Lipman, M. (1988). Critical thinking - What can it be? *Educational Leadership*, *46*(1), 38-43.

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences*, *15*(4), 431-449.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231-1243.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098-2109. https://doi.org/10.1037//0022-3514.37.11.2098

Lorencová, H., Jarošová, E., Avgitidou, S., & Dimitriadou, C. (2019). Critical thinking practices in teacher education programmes: A systematic review. *Studies in Higher Education*, *44*(5), 844-859. https://doi.org/10.1080/03075079.2019.1586331

Lowenthal, P. (2012). *Social presence: What is it? How do we measure it?* University of Colorado]. Colorado.

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), 1-13.

Markovits, H., & Brunet, M.-L. (2012). Priming divergent thinking promotes logical reasoning in 6- to 8-year olds: But more for high than low SES students. *Journal of Cognitive Psychology*, *24*(8), 991-1001.

Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology*, *100*(2), 291-309.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory and Cognition*, *39*(3), 462-476. https://doi.org/10.3758/s13421-010-0035-2

McConney, A., Woods-Mcconney, A., Schibeci, R., Maor, D., & Oliver, M. C. (2014). Inquiry, engagement, and literacy in science: A retrospective, cross-national analysis using PISA 2006. *Science Education*, *98*(6), 963-980. https://doi.org/10.1002/sce.21135

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher*, *19*(4), 10-12.

Mehta, S. R., & Al-Mahrooqi, R. (2015). Can thinking be taught? Linking critical thinking and writing in an EFL context. *RELC Journal*, *46*(1), 23-36.

Moore, T. (2004). The critical thinking debate: How general are general thinking skills? *Higher Education Research and Development*, *23*(1), 3-18.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, *24*(2), 257-271. https://doi.org/10.1080/09658211.2014.1001992

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129-140. https://doi.org/10.1177/2372732215600886

Muis, K. R., Psaradellis, C., Chevrier, M., Di Leo, I., & Lajoie, S. P. (2016). Learning by preparing to teach: Fostering self-regulatory processes and achievement during complex mathematics problem solving. *Journal of Educational Psychology*, *108*(4), 474-492.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1-8. https://doi.org/10.1038/s41562-016-0021

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142-1150.

Nickerson, R. S. (1988). On Improving Thinking Through Instruction. *Review of Research in Education*, *15*(1), 3-57. https://doi.org/10.3102/0091732X015001003

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220.

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.

Nosek, B. A. (2017). Opening Science. In *Open: The Philosophy and Practices that are Revolutionizing Education and Science* (pp. 89-99). Ubiquity Press. https://doi.org/10.5334/bbc.g

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137-141. https://doi.org/10.1027/1864-9335/a000192

Nursa'adah, E., Sari, D., & Affifah, I. (2021). Domain-specific critical thinking skills: Designing and analysis the instrument test on thermochemistry. *Journal of Physics: Conference Series*, *1731*, 012025. https://doi.org/10.1088/1742-6596/1731/1/012025

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, *5*, 114. https://doi.org/10.3389/frobt.2018.00114

Olms, C., Jakstat, H. A., & Haak, R. (2017). Elaborative feedback in clinical dental teaching—An assessment from the students' perspective. *Annals of Anatomy*, *212*, 48-54. https://doi.org/https://doi.org/10.1016/j.aanat.2017.02.007

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434.

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

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

Paas, F., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737-743.

Pariser, E. (2012). *The filter bubble: What the Internet is hiding from you*. Penguin Books.

Pellegrino, J. W., & Hilton, M. L. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. In: Committee on Defining Deeper Learning and 21st Century Skills. Division on Behavioral and Social Sciences and Education. National Research Council. Washington, DC: National Research Council.

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90-100. https://doi.org/10.1080/00461520.2021.1895796

Qualtrics. (2017). *Qualtrics*. In http://www.qualtrics.com

Quatrociocchi, W., Scala, A., & Sunstein, C. (2016). *Echo chambers on Facebook*. SSRN. Retrieved April 6 from https://ssrn.com/abstract=2795110

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin*, *144*(8), 779-796. https://doi.org/10.1037/bul0000146

Rassin, E. G. C. (2008). Individual differences in the susceptibility to confirmation bias. *Netherlands Journal of Psychology*, *64*(2), 87-93. https://doi.org/10.1007/BF03076410

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, *25*(4), 523-548.

Reeves, A. N. (2014). Written in black & white: Exploring confirmation bias in racialized perceptions of writing skills. *Nextions Yellow Paper Series*.

Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction*, *5*(1), 21-36.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1-29. https://doi.org/10.1207/s15516709cog2101_1

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529-556.

Renkl, A., & Atkinson, R. (2010). Learning from worked-out examples and problem solving. In *Cognitive Load Theory* (pp. 89-108). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511844744.007

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM : Mathematics Education*, *49*(4), 599-611. https://doi.org/10.1007/s11858-017-0834-z

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Roelle, J., & Nückles, M. (2019). Generative learning versus retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology*, *111*(8), 1341-1361. https://doi.org/10.1037/edu0000345

Roelle, J., Rahimkhani-Sagvand, N., & Berthold, K. (2017). Detrimental effects of immediate explanation feedback. *European Journal of Psychology of Education: A Journal of Education and Development*, *32*(3), 367-384.

Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. *Instructional Science*, *42*(3), 327-351.

Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science*, *36*(4), 321-350.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16-20. https://doi.org/10.1007/BF02322211

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90-100. https://doi.org/10.1037/a0015108

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior*, *27*(1), 1-18.

Schwind, C., Buder, J., Cress, U., & Hesse, F. W. (2012). Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education*, *58*(2), 787-796.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*(2), 258-266. https://doi.org/10.1016/j.chb.2008.12.011

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, *30*(9), 1371-1379. https://doi.org/10.1177/0956797619861429

Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 294-306. https://doi.org/10.3758/CABN.4.3.294

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189.

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1999). *Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess* Proceedings of the ACM symposium on virtual reality software and technology,

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*(11), 1202-1212.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. In *The Wiley Blackwell handbook of judgment and decision making, II* (pp. 924-951). John Wiley & Sons, Ltd : Chichester, UK. https://doi.org/10.1002/9781118468333

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*(3), 225-247. https://doi.org/10.1080/13546780600780796

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672-695. https://doi.org/10.1037/0022-3514.94.4.672

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: toward a test of rational thinking*. The MIT Press.

Sternberg, R. J., & Halpern, D. F. (2020). *Critical thinking in psychology* (Second edition. ed.). Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2), 165-169.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59-89.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 Years later. *Educational Psychology Review*, *31*(2), 261- 292.

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*(3), 497-510. https://doi.org/10.1037//0022-0663.91.3.497

Ten Dam, G., & Volman, M. (2004). Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction*, *14*(4), 359-379.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports*, *20*(2), 663-678. https://doi.org/10.2466/pr0.1967.20.2.663

Thayer, R. E. (1978). Factor analytic and reliability studies on the activation-deactivation adjective check list. *Psychological Reports*, *42*(3), 747-756. https://doi.org/10.2466/pr0.1978.42.3.747

Thayer, R. E. (1986). Activation-deactivation adjective check list: Current overview and structural analysis. *Psychological Reports*, *58*(2), 607-614. https://doi.org/10.2466/pr0.1986.58.2.607

Timmermans, A. C., Kuyper, H., & Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, *85*(4), 459-478.

Timmermans, A. C., Rubie-Davies, C. M., & Rjosk, C. (2018). Pygmalion's 50th anniversary: the state of the art in teacher expectation research. *Educational Research and Evaluation*, *24*(3-5), 91-98.

Tiruneh, D. T., De Cock, M., & Elen, J. (2018). Designing learning environments for critical thinking: Examining effective instructional approaches. *International Journal of Science and Mathematics Education*, *16*(6), 1065-1089. https://doi.org/10.1007/s10763-017-9829-z

Tiruneh, D. T., Gu, X., De Cock, M., & Elen, J. (2018). Systematic design of domain-specific instruction on near and far transfer of critical thinking skills. *International Journal of Educational Research*, *87*, 1-11.

Tiruneh, D. T., Weldeslassie, A. G., Kassa, A., Tefera, Z., De Cock, M., & Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development*, *64*(3), 481-505. https://doi.org/10.1007/s11423-015-9417-2

Torringa, H. (2019, May 10). Je zal Annie maar zijn. https://heleentorringa.nl/je-zal-annie-maar-zijn/

Tu, C.-H. (2002). The measurement of social presence in an online learning environment. *International Journal on E-Learning*, *1*(2).

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

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293-315.

UNESCO (2018). *Journalism, fake news & disinformation: Handbook for journalism, education and training* (C. Ireton & J. Posetti, Eds.). UNESCO.

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2020). 'Consider the Opposite' – Effects of elaborative feedback and correct answer feedback on reducing confirmation bias – A pre-registered study. *Contemporary Educational Psychology*, 101844. https://doi.org/https://doi.org/10.1016/j.cedpsych.2020.101844

Van Brussel, S., Timmermans, M. C. L., Verkoeijen, P. P. J. L., & Paas, F. (2021). Teaching on video as an instructional strategy to reduce confirmation bias—a pre-registered study. *Instructional Science*, 49(4), 475–496. https://doi.org/10.1007/s11251-021-09547-4

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*(2), 497-527.

Van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, *58*(1), 263-272. https://doi.org/https://doi.org/10.1016/j.compedu.2011.07.020

Van der Kleij, F. M., Feskens, R., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, *85*, 1-37. https://doi.org/10.3102/0034654314564881

Van der Zanden, P. J. A. C., Denessen, E., Cillessen, A. H. N., & Meijer, P. C. (2018). Domains and predictors of first-year student success: A systematic review. *Educational Research Review*, *23*, 57-77.

Van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*(2), 265-289. https://doi.org/10.1007/s10648-015-9297-3

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212-218. https://doi.org/10.1016/j.cedpsych.2010.10.004

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155-174.

Van Peppen, L. M. (2020). *Fostering critical thinking: Generative processing strategies to learn to avoid bias in reasoning* Erasmus University Rotterdam]. Rotterdam.

Van Peppen, L. M., Verkoeijen, P. P. J. L., Heijltjes, A. E. G., Janssen, E. M., Koopmans, D., & Van Gog, T. A. J. M. (2018). Effects of self-explaining on learning and transfer of critical thinking skills. *Frontiers in Education*, *3:100*. http://repub.eur.nl/pub/112157

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2018). Teachers' high-stakes decision making. How teaching approaches affect rational and intuitive data collection. *Teaching and Teacher Education*, *71*, 108-119. https://doi.org/10.1016/j.tate.2017.12.011

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2020). Sorting pupils into their next educational track: How strongly do teachers rely on data-based or intuitive processes when they make the transition decision? *Studies in Educational Evaluation*, 100865. https://doi.org/10.1016/j.stueduc.2020.100865

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, *15*(6), 589-602.

Vydiswaran, V. G. V., Zhai, C., Roth, D., & Pirolli, P. (2015). Overcoming bias to learn about controversial topics. *Journal of the Association for Information Science and Technology*, *66*(8), 1655-1672.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1-39.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129-140.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273-281. https://doi.org/10.1080/14640746808400161

Weidlich, J., & Bastiaens, T. J. (2017). Explaining social presence and the quality of online learning with the SIPS model. *Computers in Human Behavior*, *72*, 479-487. https://doi.org/10.1016/j.chb.2017.03.016

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, 1832.

Wijnia, L., Loyens, S. M. M., Derous, E., & Schmidt, H. G. (2014). Do students' topic interest and tutors' instructional style matter in problem-based learning? *Journal of Educational Psychology*, *106*(4), 919-933.

Willingham, D. T. (2007). Critical thinking - Why is it so hard to teach? *American Educator*, *31*(2), 8-19.

Yang, Y. T., & Chou, H. A. (2008). Beyond critical thinking skills: Investigating the relationship between critical thinking skills and dispositions through different online instructional strategies. *British Journal of Educational Technology*, *39*(4), 666-684.

Zimmer, F., Scheibe, K., Stock, M., & Stock, W. G. (2019). Fake news in social media: Bad algorithms or biased users? *Journal of Information Science Theory and Practice*, *7*(2), 40-53.

Zohar, A., & Schwartzer, N. (2005). Assessing teachers' pedagogical knowledge in the context of teaching higher-order thinking. *International Journal of Science Education*, *27*(13), 1595-1620.

Dankwoord (Acknowledgements in Dutch)

“Success is no accident.

It is hard work, perseverance, learning, studying, sacrifice

and most of all, love of what you are doing or learning to do”

Edson Arantes do Nascimento (Pelé)

Deze tekst hangt al heel wat jaren op mijn werkkamer en inspireert me om door te zetten. Dit proefschrift is af en een nieuwe fase ligt voor me. Uitgeleerd ben ik nooit, daarvoor ben ik veel te nieuwsgierig. Maar ik kan niet nieuwsgierig zijn zonder de ander. Ook dit proefschrift was er niet gekomen zonder anderen. Te veel om ieder bij naam te noemen, maar ik zal een poging doen.

Allereerst uiteraard mijn geweldige begeleiders: prof. dr. Fred Paas, dr. Miranda Timmermans en dr. Peter Verkoeijen. Drie toppers, ik had me geen beter team kunnen wensen.

Bedankt veur dieês zes schoên joar, ich voông ut fein um met euch te moôge samewerreke!

Fred, op de internationale en nationale congressen die ik heb bezocht, kwamen mensen uit alle windstreken bij mij langs omdat jouw naam als co-auteur bij de presentatie stond. Dat zegt veel over jouw indrukwekkende wetenschappelijke reputatie en het vervult me met trots dat ik dit traject onder jouw supervisie heb mogen doen. Je reageerde altijd snel op mijn vragen en stukken die ik voorlegde voor feedback. Je hebt mijn denken blijvend aangescherpt. Dankzij jouw begeleiding is de kwaliteit van dit proefschrift hoger dan ik vooraf ooit van mezelf verwacht had! Ook al was je in Australië aan het werk, ik kon altijd op je rekenen. Bedankt voor je inspanningen om mij te begeleiden en voor het vertrouwen dat je in me had.

Miranda, jij startte in 2014 als lector bij de pabo van Avans en we spraken elkaar vooraf al een keer. Ik wist toen nog niet dat jij mijn copromotor zou worden, maar de klik was er meteen! Jij bent in staat om scherpe vragen te stellen en mij uit te dagen met een prikkelende stelling. Steeds bleef je de verbinding met het opleiden van leraren leggen en je accepteerde ook dat ik dat soms wat minder deed. Als je agenda het toeliet, was je bij mijn presentaties. Tijdens de EAPRIL in 2015 In Luxemburg bij mijn eerste round table in het Engels heb je me gesteund en gaf je me vertrouwen. Jouw persoonlijke betrokkenheid was voor mij goud waard en heeft mede geleid tot de daadwerkelijke afronding van dit grote project. Uiteraard ook veel dank voor de financiering van de onderzoeksplaats tussen 2015 en 2021 in jouw lectoraat Leerkracht.

Peter, in het begin van dit traject troffen we elkaar vaak op vrijdagmiddag op het schoolplein bij het ophalen van onze kinderen die van dezelfde leeftijd zijn. Dat lijkt alweer zo lang geleden. In de loop der jaren werd dat schoolplein ingewisseld voor een terrasje op de Ginnekenmarkt en tijdens de lockdown aan de eettafel bij één van ons thuis. Natuurlijk spraken we over het onderzoek, maar omdat we in dezelfde levensfase zitten ook over de uitdagingen van een gezin met tieners in combinatie met werk, verbouwingen en fietsen. Je trakteerde op publicatiemuffins en proefschriftbitterballen om samen de mijlpalen te vieren. Geweldig dat je daar aandacht aan schonk! Je was er ook met wijze woorden om de balans te houden als ik het soms zwaar had. Niet voor niets hebben Lara en ik je genomineerd als supervisor of the year omdat je oog hebt voor het welzijn van je kandidaten. Wij vonden het dan ook zeer terecht dat je deze award in de wacht sleepte! Je bent deskundig, authentiek en oprecht geïnteresseerd en hebt mij met humor, metaforen, voorbeelden en oneindig veel geduld de statistiek uitgelegd en ook steeds aan het denken gezet! Wat heb ik veel van je geleerd, dank je wel. Ik hoop dat ik je nog lang mijn collega mag noemen en dat we onze samenwerking voortzetten binnen Avans Hogeschool.

De leden van de promotiecommissie, prof. dr. Lidia Arends, prof. dr. Tamara van Gog en prof. dr. Jan van Tartwijk ben ik veel dank verschuldigd voor het kritisch lezen en beoordelen van dit proefschrift en het deelnemen aan de oppositie. Prof. dr. Guus Smeets, prof. dr. Sofie Loyens en dr. Anita Heijltjes, de overige leden van de promotiecommissie, dank ik hartelijk voor hun bereidheid om met mij van gedachten te wisselen over de inhoud van mijn proefschrift. Het is een eer om ten overstaande van jullie mijn proefschrift en de stellingen te mogen verdedigen.

Hoewel ik tijdens dit PhD-traject als buitenpromovendus veel solistisch aan het werk was, is het proefschrift mede tot stand gekomen dankzij mijn collega's in de brede zin van het woord.

Het College van Bestuur van Avans Hogeschool dank ik voor het faciliteren van tijd die ik aan de onderzoeken kon besteden. In het bijzonder dank ik voormalig lid dr. Diederik Zijderveld voor zijn persoonlijke inspanning om in 2015 een doorstart van mijn promotieonderzoek mogelijk te maken.

Zoals eerder genoemd, dank ik lector dr. Miranda Timmermans en de directie van de pabo van Avans Hogeschool voor de onderzoeksplaats in het lectoraat Leerkracht. Ook lector dr. Peter Verkoeijen van het lectoraat Brein & Leren van Avans Hogeschool dank ik voor het ondersteunen van het onderzoek nadat het lectoraat bij de pabo afliep.

Veel dank aan alle collega's en oud-collega's van de pabo. Mijn onderzoekswerkzaamheden waren niet altijd zichtbaar of tastbaar, maar de blijken van oprechte interesse, peptalks en de betrokkenheid hebben me altijd zeer goed gedaan! Jullie collegialiteit tijdens het soms eenzame onderzoekswerk was fantastisch. Een aantal collega's wil ik in het bijzonder noemen. Theo, jij belde me 10 jaar geleden rond Kerstmis met de vraag: "Is promoveren iets voor jou?". En ook al is het project dat je onder mijn aandacht bracht niet het uiteindelijk proefschrift geworden, dank dat je mij uitdaagde over de grenzen van mijn mogelijkheden te denken. Mathieu, dank je wel voor je hulp bij de experimenten en de vrijmibo biertjes! Je bent vast tureluurs geworden van al het leeswerk. Nicole, bedankt voor je hulp tijdens het onderzoek en onze samenwerking op de pabo. Sabine, onze samenwerking op het gebied van onderzoek binnen de pabo is voor mij heel inspirerend. Marielle H., Mayke, Nicole, Marielle R., Esther, Marlies, Marieke, Nicole, Stephanie, Liane, Tatia, Sabine, Theo, Miranda et al., dank voor het sparren tijdens de kenniskringbijeenkomsten met het lectoraat Leerkracht. Esther, Milou, Mariola, Johan, Jurjen en Richard van het lectoraat Digitale Didactiek, ik kijk uit naar de eerste resultaten van jullie onderzoeken maar de energie in de groep is nu al significant! Dominique, dank voor je inspiratie en de tijd die je nam om te sparren. Rick, dank voor de mogelijkheden die je me bood om dit proefschrift te kunnen afronden en voor je persoonlijke betrokkenheid. Hanneke, jij hebt mij in contact gebracht met Fred, Peter en Miranda. Ik ben je er nog steeds dankbaar voor en hopelijk gaan we in het nieuwe cluster weer samenwerken.

De afwisseling tussen het onderzoek en het docentschap houdt het werk interessant. De vakgroep Nederlands en taaldidactiek blijf ik een warm hart toedragen ook al zijn mijn werkzaamheden binnen de pabo verschoven. Dank aan mijn toetsmaten van het eerste uur, Buket, Patrick en Mathieu! Mijn collega's van jaarteam 3-4, de onderwijs- en de examencommissie voor het streven naar de beste kwaliteit van onze opleiding. Dat ons dat lukt, bleek wel uit de positieve accreditatie! Melanie, het was zo tof om je te begeleiden bij jouw onderzoek. Je mag ontzettend trots zijn op de impact van jouw onderzoek op het curriculum van de pabo! Ilse, dank je wel voor al je ondersteuning in de voorbereidingen van congressen en presentaties. Monique, jouw positieve energie en tomeloze humor neem ik de rest van mijn leven mee. En dank aan alle studenten die hebben deelgenomen aan de experimenten en alle studenten die ik in de afgelopen jaren heb mogen begeleiden op de pabo en in de basisscholen. Het is tof om velen van jullie terug te zien als professionals op de basisscholen, als collega en als leerkracht van mijn eigen kinderen!

In mijn onderzoeksvoorstel uit 2015 staat onder andere de ambitie om op de onderwijsconferenties EAPRIL, EARLI, AERA, ORD en Velon mijn onderzoek te presenteren. Die ambitie heb ik waar kunnen maken en het was me een waar genoegen om naar Luxemburg, Rotterdam, Breda, Nijmegen, Tampere (Finland) en Toronto (Canada) af te reizen. Het was gaaf geweest om ook Leeuwarden, Tübingen (Duitsland) en Göteborg (Zweden) in het echt in plaats van online te kunnen bezoeken, maar dat houd ik nog graag tegoed na de coronapandemie. Ik heb op al deze en andere onderzoeksevenementen een aantal fijne mensen ontmoet en samengewerkt aan interessante symposia. Dank voor jullie inspiratie en de gezelligheid: Jeroen, Vincent, Lara, Eva, Milou, Marloes, Anita, Leonie, Andreas en Mieke. Lara, een extra woord van dank aan jou vanwege de tijd die je nam om me wegwijs te maken in het Erasmus Behavioral Lab, te sparren over het onderzoek en het uitwisselen van taken voor de experimenten.

Ik heb een aantal vrienden en vriendinnen die ik al heel lang ken: Nienke, Ria, Jeske, Kim, Ester, Nicole, Mieke en Ruud (merci voor de Limburgse vertaling in dit dankwoord). Ik ben dankbaar voor jullie vriendschap en ik verheug me op nieuwe wandelingen, saunadagjes en etentjes. Jeske, Nienke en Kim, bedankt voor jullie deskundige feedback op de Nederlandse samenvatting. Letterenmeisjes zijn de allermooiste meisjes die er zijn! Tijdens de basisschooljaren van onze kinderen en vanuit de Strijenlaan is daar nog een club bijgekomen: Mirjam en Daan, Marijke en Ludolf, Karlijn en Jeroen, Liesbeth en Bob, Geertje, Carola en Laura: het volgende rondje is van mij! Lieve Karin, vaak heel dichtbij en soms van ver als je op reis bent met Berry, jij bent er altijd voor mij.

Kim en Jeske, vanaf de start van dit traject heb ik mij voorgesteld dat jullie naast mij staan als paranimfen. Het geeft mij een geweldig gevoel dat het echt zo ver is gekomen! Tussen jullie in sta ik daar stevig. Bedankt dat jullie mij altijd het voordeel van de twijfel geven, voor de oermoedertips en ik hoop dat we samen nog lang kunnen genieten van een zuurtje, een bittertje, een zoutje maar vooral van het zoete in het leven.

Mijn schoonfamilie Gerda, Hans, Kristel, Freek, Julia en Taeke dank ik voor jullie geduld en interesse in het proefschrift. Mijn ouders dank ik voor de kansen die ik mocht grijpen om me te ontwikkelen en natuurlijk voor alle oppashulp.

Tot slot mijn dierbaarste dierbaren. Sander, Faas en Sarie, het boek is af! Het is ontzettend fijn om dit moment met jullie te vieren. Lieve Sander, het was een bumpy road maar we hebben het doorstaan en we zijn pas op de helft van ons leven! Er liggen mooie dingen in het verschiet en ik zou ze met niemand anders willen beleven dan met jou. Bedankt voor je geduld, the floor is yours now. Lieve Faas en Sarie, jullie zijn het allermooiste dat me ooit is overkomen! Jullie zijn mijn onthaasters geweest tijdens afgelopen jaren en laten mij telkens zien wat er werkelijk toe doet in het leven. Het is zo bijzonder om jullie te zien opgroeien. En hoe graag ik jullie ook iets vertel over wetenschappelijk onderzoek met mijn juffenstem, blijf zelf nadenken én voelen wat bij jou past. Vertrouw op je eigen kompas, volg je eigen pad en maak je dromen waar want dat maakt het leven oneindig veel leuker. Als je het even niet meer weet, dan is er altijd de infinite love van Susafarie want ik hou dus echt oneindig veel van jullie!

Suzan, januari 2022

Zin

als je ergens
een punt achter zet
ga je verder
met een hoofdletter.

Merel Morre

“Dit is waarschijnlijk het geheim: denken kwam er volstrekt niet aan te pas, enkel was er altijd maar en zonder aarzeling de dringende gewaarwording dat, wat ze je ook proberen wijs te maken, dingen niet noodzakelijkerwijze zijn zoals je wordt geleerd.

Er is zoveel meer mogelijk. Altijd en op elk gebied. Wij moeten het daarom wel zonder na te denken doen, dat ene wat zij geen van allen in ons zien, dat ongekende.

Die ene keer, die geen ruimte in ons laat voor twijfel, mag een mens niet afgaan op een afweging. Anders zouden we aan wat het meest de moeite waard is nooit beginnen, ons eigen pad bijvoorbeeld of de liefde, het loskomen, de lucht. ‘De mens…’ riep hij en stak zijn arm op. ‘Vliegt!’ ”

Uit: De gevleugelde – Arthur Japin

“Iemand die mateloos nadenkt heeft waarschijnlijk een grotere angst voor het leven dan iemand anders.”

Uit De Vriendschap - Connie Palmen