



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

Suzan van Brussel¹ · Miranda Timmermans¹ · Peter Verkoeijen^{2,3} · Fred Paas^{3,4}

Received: 19 April 2021 / Accepted: 23 January 2023 / Published online: 28 February 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

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, preparing to teach, or re-study (control condition), and as a post-test 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 significantly more accurate and complete open-minded lessons than 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

✉ Suzan van Brussel
sjm.vanbrussel@avans.nl

¹ Primary Education Institution, Avans University of Applied Sciences, PO Box 90116, 4800 RA Breda, The Netherlands

² Learning and Innovation Centre, Avans University of Applied Sciences, Breda, The Netherlands

³ Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

⁴ School of Education/Early Start, University of Wollongong, Wollongong, Australia

Introduction

Learners often rely on superficial and inefficient learning strategies that target rote learning, and that do not lead to meaningful knowledge. Meaningful knowledge is an interrelated collection of new and existing knowledge about a particular topic which is necessary to understand the learning material in the long term, and which leads to new knowledge structures (Dunlosky et al., 2013; Fiorella & Mayer, 2013, 2016). It is therefore considered important to encourage learners to process learning material through generative learning strategies. If learning is generative, learners try to make sense of the instructional material presented to them (Fiorella & Mayer, 2016). It is an interplay between selecting, organising, and integrating new and prior information. Hence, generative learning strategies contribute to effective learning because learners are actively involved in making the to-be-learned information meaningful by paying attention to important aspects of the new information, by reorganising 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 et al., 2019a, 2019b; Kobayashi, 2019; Renkl, 1997; Roscoe & Chi, 2008). Preparing to teach means that learners study the learning material by preparing a lesson with a teaching expectancy (Kobayashi, 2019; Muis et al., 2016). This implies that learners study the material and prepare an actual lesson on paper while keeping in mind that they have to explain it at a later moment to someone 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, organise 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 et al., 2019a, 2019b). 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, 2019a, 2019b). 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 re-studying the learning material (Fiorella & Mayer, 2013).

Learning by teaching on video has been shown to be a beneficial strategy for acquiring procedural knowledge e.g., to learn to reduce the confirmation bias through three “consider the opposite”-stages (Van Brussel et al., 2021), and acquiring problem-solving skills from step-based worked examples (Hoogerheide et al., 2019a). It is an open question as to whether learning by teaching on video supports student teachers’ learning through authentic tasks such as preparing open-minded citizenship lessons. There is no reason to believe

that teaching on video does not generalize to authentic tasks. However, this generalization question has not yet been addressed. For student teachers, preparing a lesson plan and afterwards teach it, is an authentic task because it simulates the task, they have to perform during their future job (Abrami et al., 2015) Simulating teaching while learning by teaching on video might therefore be seen as an additional meaningful learning opportunity for student teachers which might enhance effective learning.

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 (Cavojoja 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 (Cavojoja et al., 2018; Nickerson, 1998; Schwind et al., 2012; Stanovich et al., 2016; Sternberg & Halpern, 2020; Tversky & Kahneman, 1974). As a result, the confirmation bias can lead to one-sidedness. Perspective taking is an important element of citizenship education to avoid this one-sidedness (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 and one-sidedness 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 these 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 education 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 study¹

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 re-study. Participants were student teachers who first completed the Actively Open-minded Thinking (AOT;

¹ The present study was pre-registered at the Open Science Framework: Hypotheses, planned data-collection, planned methods, planned analyses and the datasets generated and analysed during the present study are available on the [Open Science Framework](#).

Stanovich & West, 2007). After one week all participants received an instruction² on what open-minded lessons, confirmation bias and perspective taking are and why it is important to gain knowledge about these subjects. 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 re-studying the text for 10 min. During the learning phase, feelings of social presence and arousal were measured through questionnaires in the teaching on video and preparing to teach condition.³ After the learning phase, all participants completed the AOT for the second time. As a post-test, all participants wrote a lesson plan for a topic within the context of citizenship education lesson in primary education (e.g., racism, obesity). Finally, all participants completed a conceptual knowledge test to assess 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 following line of reasoning: Participants who learn through teaching on video often experience higher feelings of social presence which induce higher arousal levels because of addressing an audience (i.e., the social presence hypothesis; Gunawardena, 1995; Hoogerheide et al., 2019a, 2019b). 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 et al., 2019a, 2019b), 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-studied the instructional content. Compared to the generative learning strategies, re-studying does not involve deep processing of the 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 hypothesize that the accuracy and completeness of the open-minded lesson plan would be highest in the TOV condition, followed by the PTT condition, in which accuracy and completeness would be higher compared to 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 to participants who only

² The instructional text in Dutch can be found on the *Open Science Framework*.

³ Due to a technical error, contrary to the pre-registration, the arousal questionnaire was not administered in the control condition.

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

Furthermore, we exploratively compared mean AOT pre-test to post-test 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 0.80 for the One-Way (single factor) ANOVA, under a significance level of 0.05, and a medium effect size ($f=0.25$), 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 ($M_{\text{age}}=21.60$, $SD=4.99$, 153 women) from six Dutch primary education 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.

By pre-registering 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 pre-registration 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 and measures 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.

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. 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, p. 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 (See Appendix B) was constructed by Van Brussel et al. (2021), and in their study the Cronbach's alpha was 0.73. In the social presence questionnaire 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 (10×1), to 50 (10×5). The higher the score, the higher the feelings of social presence.

Arousal

To measure arousal, the activation-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 0.82 on "energy", and 0.82 on "tiredness", but questionable Cronbach's alphas on both "tension" (0.67) and "calmness" (0.60).

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

Table 1 Overview of the measures and objectives

Measure	We used this measure to assess whether
Actively Open-minded Thinking Test	The three conditions would differ in mean open-mindedness one week before and immediately after the intervention
Quality of the explanation	The TOV and PTT condition would differ in mean completeness and mean accuracy of the explanation instructional content (i.e., content related to the confirmation bias)
Social presence	The TOV and the PTT condition would differ in mean self-reported social presence
Arousal	The TOV and the PTT condition would differ in mean self-reported arousal
Conceptual knowledge	All three conditions would differ in the knowledge that was addressed in the instruction after the intervention
Quality of the lesson plan (final test)	All three conditions would differ in the final test on the mean completeness and mean accuracy of a 'confirmation bias free' lesson of a citizenship education lesson

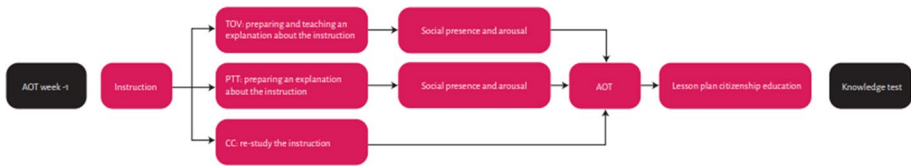


Fig. 1 Procedure of the experiment

first author and was aimed at testing participants' knowledge about the concepts that were addressed in the instruction (See Appendix A).

All measures and objectives of the measures that are used in the present study are presented in Table 1.

Procedure

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

One week before the experiment, participants completed the AOT for the first time. After 1 week, participants received an e-mail with the link that led to the Qualtrics platform. Participants read the assignment and its goals 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 without any further instructions on how to process the information afterwards. 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. 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.

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 later. 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.⁴ The recording of the explanation was sent by e-mail or WhatsApp to the experimenter. In the preparing to teach group (PTT), participants only wrote a preparation of their explanation for their peers in Qualtrics. Their instruction was as follows: “Prepare an explanation of what you have just learned for your peer student teachers who are not participating in this project. They will read your explanation later. This information is also important for your peer students to learn to think critically. Therefore, make sure your explanation is accurate and complete. Start typing your explanation in the text box below.” Participants in the TOV and PTT conditions were called upon to set the time for this phase for 10 min. In the re-study control condition (CC), participants re-studied the instructional text for 10 min after which they were automatically forwarded to the next page in Qualtrics. The re-study instruction was: “You now have the opportunity to study the subject matter again before continuing with assignments on this subject. You have ten minutes. When the time is up, the program will automatically proceed to the next page. You are not allowed to take notes. At the bottom of the page, you can see how much time you have left.” Then these participants had to indicate how often they re-studied the instructional text: once, twice, or otherwise, showing how often they studied the instructional text.

After this, the 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- and 12-year-olds) 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, LGBTQ+ community, religion / belief, childhood obesity, mouth masks in public transport⁵ and (black) Pete. Please note, there is a minimum number of characters that you must use. If you cannot click to the following page yet, you will have to explain more.” In their lesson plan, participants were supposed to show open-mindedness by taking perspective on the chosen topic through considering opposites and alternatives to their own point of view without giving the explicit instruction to do so. The

⁴ 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.

⁵ 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.+

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 and results

Based on our data analysis plan of the pre-registration, we checked variables on missing data and outliers. There were 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, the conceptual knowledge test and the social presence and arousal questionnaires, a minimal Cronbach's alpha level of 0.70 was set as a threshold for the analysis of a sum score or average score. In all analyses below, a significance level of 0.05 was used as a threshold for statistical significance. Eta-squared (η^2) is reported as measure of effect size for the ANOVAs for which 0.01 is considered small, 0.06 medium, and 0.14 large.

We asked TOV and PTT participants to set the time for 10 min in which they have to prepare and teach (TOV) or prepare (PTT). Qualtrics automatically reported when TOV and PTT participants clicked through to the following page (namely the social presence questionnaire). In the control condition, the page automatically proceeded after 10 min. However, the actual time spent on reading and learning to prepare (PTT) or reading and learning through teaching video (TOV) somewhat deviated from the suggested time. We observed five outliers at the upper side of the reading time for the instructional text. Without these outliers, the mean time spent on reading the instruction was 9.45 min ($SD = 11.78$). TOV participants spent on average 37.05 min ($SD = 116.64$) to prepare and teach. PTT participants spent on average 8.78 min ($SD = 3.97$) to prepare. Analyses without the time-on-task outliers in the TOV condition on dependent variables did not influence the results, therefore we decided to include the data of all participants. A One-Way ANOVA on the time-on-task did not reveal a significant difference between the means of the TOV and PTT conditions: $F(2,102) = 0.645$, $p = 0.527$. In addition, we analysed the correlation between the time-on-task of the preparation phase collapsed across the TOV and PTT conditions and the final test results. The results show no significant correlation between the time-on-task and the accuracy of the lesson plan, $r(98) = 0.165$, $p = 0.101$, nor on the completeness of the lesson plan, $r(98) = 0.155$, $p = 0.124$. Furthermore, the correlation between the time-on-task and the results of the conceptual knowledge test was also not significant, $r(105) = 0.085$, $p = 0.38$. Our results show that the time-on-task differences between the TOV and PTT condition were due to outliers: without these outliers the time-on-task means were similar for the TOV and PTT conditions. Moreover, the mean time-on-task was comparable to the time allotted to participants in the re-study condition. 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). Furthermore, and crucially, the outcomes of our correlation analysis show that time-on-task is not a confound in the current study.

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 0.53. The test was reliable on the second measure: Cronbach's alpha = 0.83. To determine whether there were differences between the three groups, we conducted a 3 (Condition) \times 2 (Pre-test, Post-test) Mixed ANOVA.

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: $\kappa=0.73$, accuracy explanation: $\kappa=0.73$, completeness lesson plan: $\kappa=0.71$, and accuracy lesson plan: $\kappa=0.79$. Therefore, one rater scored the remainder of the texts and these results were used in the analyses.

The quality of the explanations and the open-mindedness of the lesson plans 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). For example, when there were missing elements in the explanation of open-mindedness, the participant earned 0.5 points. The approach for this analysis was based on Hoogerheide et al., (2019a, 2019b) 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 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 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 ADAFL 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 0.83. After deleting the first item, the Energetic subscale showed a Cronbach’s alpha of 0.76. However, the other two subscales were not reliable: For Tiredness, even after deleting item 3, Cronbach’s alpha was 0.65, and for Calmness the Cronbach’s alpha was only 0.61 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 Tension and Energetic subscales and the individual items of the subscales Tiredness and Calmness, we conducted independent samples t -tests to determine differences between conditions.

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=0.41. Therefore, contrary to what was pre-registered, we conducted explorative One-Way ANOVAs on the individual items.

Results for open-mindedness

We conducted a 3 (Condition: TOV, PTT, CC) \times 2 (Pre-test vs. Post-test) Mixed ANOVA with Condition as a between-subjects factor on the AOT scores. The initial measurement had a low Cronbach’s alpha of 0.53. The test was reliable on the second measure: Cronbach’s alpha=0.83. In Table 2, the mean individual scores and SD’s of both test moments are presented. We found a main effect of Test Moment: $F(1,173)=3515.33$, $p<0.001$, $\eta^2=0.95$: All participants scored higher on the second measurement compared to the first.

Table 2 Mean individual item scores and standard deviations of the actively open-minded thinking tests per condition

	AOT 1	AOT 2
TOV $n=51$	2.93 (.17)	4.42 (.37)
PTT $n=54$	2.92 (.20)	4.44 (.37)
CC $n=76$	2.93 (.18)	4.45 (.35)
Total mean	2.93 (.18)	4.44 (.36)

Table 3 Mean scores and standard deviations, per condition for the explanation and the lesson plan

	Explanation		Lesson Plan	
	Completeness (Max 10 points)	Accuracy	Completeness (max 4 points)	Accuracy
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)

However, we found no main effect of Condition: $F(2,173)=0.049$, $p=0.952$. Also, no interaction effect was found; $F(2, 173)=0.19$, $p=0.824$ and $\eta^2=0.002$.

Results for the quality of 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 post-test. See Table 3 for the relevant descriptive statistics.

The explanation (TOV and PTT)

The conditions neither differed significantly on completeness: $F(1,103)=0.71$, $p=0.40$ with a small effect size of $\eta^2=0.007$, nor on accuracy: $F(1,103)=0.12$, $p=0.73$, again with a small effect size of $\eta^2=0.001$. Two outliers for accuracy were detected but running the analyses without them did not yield other results: $F(1,101)=0.049$, $p=0.825$, $\eta^2=0.001$.

The lesson plan (TOV, PTT and CC)

All participants wrote a lesson plan as a post-test and there were no outliers. Participants chose one of the given topics for the lesson plan: Radicalization of young people (1.14%), Migration and refugees (5.47%), Religion (6.31%), Childhood obesity (12.36%), LGBTQ+community (21.06%), Mouth masks in public transport (22.33%), and Black Pete (31.35%). 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, $F(2,173)=3.32$, $p=0.039$, $\eta^2=0.037$, and for accuracy, $F(2,173)=5.05$,

Table 4 Scores of overall quality of the lesson plan per condition as percentages per scale

	Condition		
	TOV (%)	PTT (%)	CC (%)
Scale	<i>n</i> = 51	<i>n</i> = 54	<i>n</i> = 71
Insufficient	20	19	21
Sufficient	33	33	34
Good	20	28	25
Excellent	27	20	20

$p=0.007$, $\eta^2=0.055$. As a follow up, a planned Helmert contrast was performed. The contrast showed that the completeness of the lesson plan in the TOV condition was not significantly lower with a contrast estimate of -0.33 ($SE=0.22$), $p=0.133$, compared to the combined completeness score of PTT and CC. A contrast estimate of -0.47 ($SE=0.24$), $p=0.050$ 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 -0.36 ($SE=0.20$), $p=0.073$ compared to the combined accuracy score of PTT and CC. A contrast estimate of -0.54 ($SE=0.22$), $p=0.014$ did reveal that PTT participants scored lower on accuracy compared to CC participants. In sum, participants who re-studied the instruction delivered a more complete and accurate lesson plan compared to participants who processed the instruction through preparing 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 ($H(2)=0.311$, $p=0.856$). See Table 4 for percentages per condition per ordinal scale: insufficient, sufficient, good, and excellent.

We analysed whether topic choice and condition were associated by conducting a Chi-square test. This test failed to reveal a significant association between the choice of topic and condition, $\chi^2(12, n=176)=5.11$, $p=0.954$. Hence, topic choice cannot explain the results we found on the dependent variables. The choice of topic does not vary between conditions and therefore does not offer an alternative explanation for the results.

Results for social presence

To achieve the minimum required Cronbach's alpha level of 0.70, items 3 and 10 were deleted, which resulted in a Cronbach's alpha of 0.70. The analysis was conducted with the subset of eight items. The mean score for the TOV participants on the subset of the social presence questionnaire was 29.24 ($SD=4.51$), and for the PTT participants 30.56 ($SD=4.51$). This difference was not significant: $t(103)=-1.50$, $p=0.137$.

Results for arousal

See Table 5 for the descriptive statistics of the ADACL subscales for the TOV and PTT condition. For the Tension subscale, after deleting item 6, the Cronbach's alpha was 0.83. After deleting the first item, the Energetic subscale showed a Cronbach's alpha

Table 5 Mean scores and standard deviations of the subscales of the ADACL

N	Tension	Energetic	Tiredness	Calmness
TOV (<i>n</i> =51)	2.96 (.81)	2.02 (.69)	3.45 (.60)	2.30 (.67)
PTT (<i>n</i> =54)	3.35 (.74)	2.21 (.63)	3.33 (.58)	2.13 (.60)

Table 6 Conceptual knowledge test scores and standard deviations per proportion correct per item per condition

		TOV	PTT	CC
		<i>n</i> =51	<i>n</i> =54	<i>n</i> =76
1	The definition of critical thinking	.28 (.39)	.19 (.39)	.28 (.38)
2	Explanation of the confirmation bias	.74 (.34)	.78 (.33)	.74 (.35)
3	A characteristic of an open-minded person	.87 (.28) ^A	.80 (.33) ^B	.68 (.39) ^A
4	Steps that contribute to preparing a confirmation bias-free explanation	.76 (.31)	.70 (.36)	.70 (.36)
5	Explain the importance to reduce the confirmation bias	.76 (.32) ^A	.82 (.33) ^A	0.51 (.39) ^B
6	Case and an adequate response	.47 (.44)	.33 (.40)	.53 (.68)

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

of 0.76. However, the other two subscales were not reliable: For Tiredness, even after deleting item 3, Cronbach's alpha was 0.65, and for Calmness the Cronbach's alpha was only 0.61 after deleting item 14. Therefore, explorative analyses of these subscales were conducted on individual items.

An independent samples *t*-test showed that for the subscale Tension, a significant difference with higher mean scores for the PTT participants was found: $t(103) = -2.541$, $p = 0.001$. No significant difference was found for Energetic $t(103) = -1.461$, $p = 0.147$. 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 ($M = 3.35$, $SD = 0.98$) compared to PTT ($M = 2.61$, $SD = 1.11$).

Results for the conceptual knowledge test

Since the test was not reliable ($\alpha = 0.41$), results are presented on item level in Table 6. We conducted explorative One-Way ANOVAs which revealed significant differences on item 3: $F(2,173) = 4.723$, $p = 0.010$, and item 5: $F(2,173) = 13.776$, $p < 0.001$. To find out which conditions differed, independent *t*-tests were conducted on these two items.

Discussion and limitations

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 formed 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 post-test, participants in the control condition designed an open-minded lesson plan that was more accurate and more 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 limitations of this 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 et al., 2019a, 2019b), or counting the number of personal references in the explanation (e.g., Jacob et al., 2021; Hoogerheide et al., 2016; Lachner et al., 2018) might have yielded other results. For follow-up research, it is interesting to use both types of measurements and to measure feelings of social presence during 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 a peer would actually watch their video. 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 et al., (2019a, 2019b), 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 post-test, 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 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 et al., (2019a, 2019b), 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 in that study also led to better scores on a post-test 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 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 post-test 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). In sum, 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.

Retrieval practice, sometimes referred to as testing, is a strategy in which learning is enhanced by asking a learner to retrieve information from memory (Agarwal & Roediger, 2018). In our study, during the practice phase, TOV and PTT were in fact combined with retrieval practice and retrieval practice typically reveals its positive effect on performance after a longer term (Rawson et al., 2013; Roediger & Karpicke, 2006, but see Van Gog et al. (2015) for contrasting findings). Retrieval practice might explain why we did not find an advantage of TOV or PTT on the main dependent variables. This is potentially consistent with other studies in the literature. For example, in the study by Hoogerheide et al., (2019a, 2019b), in which the TOV participants had a worked example at hand during teaching, TOV scored better on a final test compared to a control condition in which the example was re-studied. Furthermore, Jacob et al. (2021), also showed that a retrieval practice control condition scored lower compared to a written or oral practice condition in which participants had the instructional materials at hand. Also, Lachner et al. (2018), did not find any differences between a retrieval practice condition and explaining conditions where participants were allowed to consult the instructional materials. The latter finding suggests that retrieval practice might be the effective ingredient in strategies that involve (self-)explanation. It might be interesting for future studies to investigate whether this is true or whether the combination of retrieval practice and (self-)explanation has an effect that is larger than the effect of any of the constituent components. However, in the short term, immediately after processing the learning material, often no differences are found between retrieval practice and a more superficial learning strategy such as re-studying (Hoogerheide, Vincent et al., 2019). 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 administered a delayed test, for example 1 week after the instruction. All in all, whether or not retrieval practice was the decisive factor and whether retrieval practice in combination with teaching on video or preparing to teach has additive effects, is still an open question and might be of interest for future research.

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 single trait that could discriminate between participants. Therefore, we have to interpret the AOT scores cautiously. Taken together, 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 is addressed in the classroom might enhance performance even more. This might be the focus of future research.

Finally, we discuss factors in our study that might limit our conclusions. One could suggest adding a retrieval practice control condition in which participants have to summarize the instructional content. However, if one uses a learning strategy that is much used by higher education students, such as summarizing, then the effect compared to TOV and PTT is probably much smaller because summarizing promotes generative learning as well for skilled summarizers in higher education (Dunlosky et al., 2013; Fiorella & Mayer, 2016). To demonstrate an effect of a generative learning strategy, one needs to make a comparison with a control condition in which a non-generative learning strategy is used. Re-studying, as our participants in the control condition did, is often used for this (e.g., Annis, 1983; Coleman et al., 1997; Fiorella & Mayer, 2013; Hoogerheide, 2016; Hoogerheide et al., 2019a, 2019b; Kobayashi, 2019; Renkl, 1997; Roscoe & Chi, 2008).

The short instruction and the ten minutes preparation time might not be an assignment that would be given in a real-world situation to teacher students. From our earlier studies (Van Brussel et al., 2020, 2021), it is shown that the confirmation bias was reduced on abstract tasks that could be solved procedurally, such as Wason's four-card selection tasks (Wason, 1960). Participants in those studies were also briefly instructed. The length of the instruction might therefore not have led to different results in the current study. In contrast with the current study, participants received feedback on the practiced tasks. In the real world, feedback is provided to student teachers on their lessons plans before teaching the lesson. Therefore, expanding the instruction and practice phase with feedback on the explanation in future studies, might shed a light on whether a higher degree of external validity leads to differences between the quality of the lessons plans in the three conditions.

Conclusion

The current experiment did not reveal any differences between both generative learning strategies (teaching-on-video and preparing to teach), and a re-study control condition on conceptual knowledge. Furthermore, re-study had a positive effect on designing an open-minded lesson at least compared to PTT. These results might in part be due to the fact that TOV and PTT involved both elaboration and retrieval practice or because participants in both conditions did not experience remarkably high levels of social presence. Hence, the results of the current experiment indicate that more research is needed to answer the theoretically and practically relevant question when TOV and PTT are most effective. For example, to examine hypotheses in which social presence and arousal are more distinctive from each other. Then mediation analyses could provide more insight

in the relationship between the treatment and the outcome measures. Lastly, we found an overall increase on mean active open-mindedness. This result might also be relevant to the educational practice as it shows that a small intervention might enhance the adaptive dispositions that teachers need for example in citizenship education lessons.

Appendix 1

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 2

Social presence measurement for the teaching on video and preparing to teach condition (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 3

Activation deactivation adjective check list (ADAACL, 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 4

Scoring forms for the explanation and lesson plan

Explanation (TOV and PTT)	Completeness	Accuracy
	Present = 1	Accurate description = 1
	Missing = 0	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*

Explanation (TOV and PTT)	Completeness Present = 1 Missing = 0	Accuracy Accurate description = 1 Partly accurate description = 0.5 Wrong description = 0
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 (maximum score = 5 × 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)		

Author contribution All authors contributed to conception and design of the present experiment. SB was primarily responsible for the material preparation, data collection and analyses. The first draft of the manuscript was written by SB and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This research did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability This study was pre-registered on the Open Science Framework: https://osf.io/3prvu/?view_only=c7d7264829fe40e5a7fa8d5bc0454684

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All subjects gave informed consent prior to participating in the present experiment.

References

- 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>
- 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. <https://doi.org/10.1016/j.lmot.2015.11.002>
- Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications, 2*(1), 39–47.
- Baehr, J. (2011). The structure of open-mindedness. *Canadian Journal of Philosophy, 41*(2), 191–213.
- 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.
- Baron, J. (2008). Actively open-minded thinking. In J. Baron (Ed.), *Thinking and deciding* (4th ed., pp. 199–232). Cambridge University Press.
- Boyle, G. J., Saklofske, D. H., & Matthews, G. (2015). *Measures of personality and social psychological constructs*. Academic Press.
- 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>
- 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*. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020104>
- 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.
- 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>
- 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.
- 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. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741.
- 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.
- 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.
- 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>
- 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. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, *41*, 80–91.
- Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & Van Gog, T. (2019a). 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. (2019b). 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>
- 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., 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>
- Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. <https://doi.org/10.1080/09658210802647009>
- 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>
- 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., Ly, K., & 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.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
- 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.
- Muis, K. R., Psaradellis, C., Chevri er, 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.
- Munaf o, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie-du-Sert, N., Uri, S., Eric-Jan, W., Jennifer, J. W., & 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. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.
- 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>
- 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>
- 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.
- Qualtrics. (2017). *Qualtrics*. In [12–2017]. <http://www.qualtrics.com>
- Rawson, K. A., Dunlosky, J., & Sciarrelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, *25*(4), 523–548.
- 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
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.
- Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, *36*(4), 321–350.
- 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.

- 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.
- 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., & 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* (2nd ed.). Cambridge University Press.
- 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>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Van Gog, T., Kester, L., Dirx, 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.
- 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.ced>
- 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*. <https://doi.org/10.1007/s11251-021-09>
- 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 [Article]. *Teaching and Teacher Education*, 71, 108–119. <https://doi.org/10.1016/j.tate.2017.12.011>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.