Effects of creating video-based modeling examples on learning and transfer

Vincent Hoogerheide*, Sofie M.M. Loyens, Tamara van Gog

Institute of Psychology, Erasmus University Rotterdam, The Netherlands

A R T I C L E   I N F O

Article history:
Received 14 October 2013
Received in revised form 25 April 2014
Accepted 27 April 2014
Available online 21 May 2014

Keywords:
Example-based learning
Modeling
Study intention
Video creation

A B S T R A C T

Two experiments investigated whether acting as a peer model for a video-based modeling example, which entails studying a text with the intention to explain it to others and then actually explaining it on video, would foster learning and transfer. In both experiments, novices were instructed to study a text, either with the intention of being able to complete a test (condition A), or being able to explain the content to others (condition B and C). Moreover, students in condition C actually had to explain the text by creating a webcam-video. In Experiment 1 (N = 76 secondary education students) there was no effect of study intention on learning (A = B), but explaining during video creation significantly fostered transfer performance (C > B; C > A). In Experiment 2 (N = 95 university students), study intention did have an effect on learning (C > A; B > A), but only actual video creation significantly fostered transfer performance (C > A).

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Example-based learning is an effective instructional strategy that has been studied from different perspectives. Research from a cognitive perspective (e.g., cognitive load theory; Sweller, 1988; Sweller, Van Merrienboer, & Paas, 1998) has mainly focused on observational learning from worked examples, which consist of a written, step-by-step worked-out procedure for completing the learning task. This is usually an “ideal” or “didactical” procedure, reflecting how a student should learn to complete a task, which may differ from how an expert would actually handle it, since experts sometimes can skip or chunk steps (Ericsson & Staszewski, 1989). Research from a social-cognitive perspective (e.g., social learning theory; Bandura, 1977, 1986; cognitive apprenticeship; Collins, Brown, & Newman, 1989) has focused on observational learning from modeling examples in which a human model or humanoid agent demonstrates and explains how to complete a task (see Van Gog & Rummel, 2010). These models sometimes demonstrate an ideal, didactical procedure for the task, but they may also display “natural” behavior, which entails making and correcting errors (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002). In modeling examples, the model can be either an adult (e.g., Schunk, 1981; Simon & Werner, 1996) or a peer student (e.g., Braaksma et al., 2002; Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a, 2013b; Schunk & Hanson, 1985).

Research inspired by the cognitive perspective has demonstrated the effectiveness and efficiency of example-based learning. For novices, instruction consisting of example study (alternated with problem solving) leads to better learning outcomes with less investment of time and mental effort than instruction consisting of problem solving only (Atkinson, Derry, Renkl, & Wortham, 2000; Paas & Van Gog, 2006; Renkl, 2011; Sweller et al., 1998; Van Gog & Rummel, 2010) and instruction consisting of tutored problem solving (Salden, Koedinger, Renkl, Alevén, & McLaren, 2010). Research inspired by the social-cognitive perspective has not only demonstrated that example-based learning can be effective for learning, but also that it can increase learners’ self-efficacy, which is the perceived belief a learner has for learning, or performing a task at a certain level (Bandura, 1997; Schunk, 1987).

As mentioned above, peer students are known to be effective modeling examples, improving learning of students who observe them (Groenendijk et al., 2013a, 2013b; Schunk, 1987). For educators, an interesting question is whether there would also be potential benefits for learning, for the peer students who act as models in the examples (i.e., for the students who explain and/or demonstrate a task). However, despite the fact that a lot of research has investigated the effects of observing modeling examples, little is known about the effects on learning and transfer that acting as a
Why is it that different comprehension monitoring processes (e.g., asking oneself:
the study intention of being able to explain to others could result in
Bargh & Schul, 1980; Benware & Deci, 1984; Renkl, 1995). Moreover,
to the more passive approach of studying to complete a test (e.g.,
dent (i.e., teaching expectancy) can invoke an active study approach,
learners to study with the expectation of teaching to another stu-
about their own capabilities.

If students have to act as a peer model for a video-based
modeling example and are not yet experts on the topic them-
seves, they first have to study learning materials on the subject.
These learning materials are studied with a different intention than
the common intention of studying for a test. That is, the materials
are studied with the intention of being able to explain the task to
others. Secondly, the peer model actually explains the task during
the creation of the video-based modeling example. Both steps may
affect students’ learning outcomes (with better outcomes being
reflected by higher retention and transfer test performance reached
with equal or less effort investment on those tests) and beliefs
about their own capabilities.

Instructing learners to study with the intention of being able to
successfully explain a task to others might invoke a more active
study approach and cause learners to focus less on absorbing new
facts and more on interpreting and integrating new knowledge
(Benware & Deci, 1984). Some studies have shown that instructing
learners to study with the expectation of teaching to another stu-
dent (i.e., teaching expectancy) can invoke an active study approach,
and enhance learning processes and/or outcomes when compared
to the more passive approach of studying to complete a test (e.g.,
Bargh & Schul, 1980; Benware & Deci, 1984; Renkl, 1995). Moreover,
the study intention of being able to explain to others could result in
different comprehension monitoring processes (e.g., asking oneself:
“Why is it that … ?”; “Do I understand … ?”; “Can I explain … ?”),
and could involve self-explanation processes, both of which have
been shown to foster deep learning and understanding (compre-
hension monitoring: Graesser, Baggett, & Williams, 1996; Sternberg,
1987; self-explaining: Chi, Bassok, Lewis, Reimann, & Glaser, 1989;
Chi, De Leeuw, Chiu, & LaVancher, 1994; Renkl, 1997, 2002; deep
questions and explanations: Craig, Gholson, Brittingham, Williams,
& Shubeck, 2012; Craig, Sullins, Witherspoon, & Gholson, 2006).
Enhanced understanding should be particularly beneficial for transfer performance (e.g.,

Next to the effect an explanation study intention might have,
actually explaining the learning materials to another (non-present)
person during the creation of a video-based modeling example
might further improve learning outcomes. It has been shown that
generating explanations can foster learning more than rereading or
receiving explanations (Lombrozo, 2012). For example, asking
learners to generate explanations can help them to identify and
then repair knowledge gaps (Chi, 2000); to integrate new knowl-
edge with prior knowledge (Chi et al., 1994; Lombrozo, 2006), and
to transform declarative knowledge into applicable procedures
(Chi et al., 1989, 1994). Whereas these studies only prompted students
to explain to themselves (i.e., self-explanation), explaining with the
intention of providing instruction that can be shared with others
(as one would do when creating a video) can be seen as what
Leinhardt (2001) refers to as providing instructional explanations.1
According to Leinhardt, providing instructional explanations differs
from simply stating or describing a concept or procedure, by more
carefully examining it. That is, providing a full explanation of the
concept or procedure in which key features are identified, con-
nections to prior knowledge are made, and effective and important
examples are provided. In other words, providing such explana-
tions would foster deeper processing and elaboration of the
learning materials, which might foster the explainer’s learning
outcomes and especially, transfer performance.

Indeed, actively providing such instructional explanations to
others during small group discussions (Van Blankenstein, Dolmans,
Van der Vleuten, & Schmidt, 2011) and during tutoring (Cohen,
Kulik, & Kulik, 1982) has been shown to aid learning. The finding
that tutoring is not only effective in terms of the tutee’s learning,
but also in terms of the tutor’s learning (Cohen et al., 1982), is
interesting because tutors also prepare by studying learning ma-
terials with the intention of being able to explain those to others,
and subsequently explain what they have learned to the tutee. The
tutor learning effect not only applies when the knowledge and age
gap between tutor and tutee is large (e.g., Juel, 1986; Sharpley,
Irwine, & Sharpely, 1983), but also when that gap is small (e.g.,
Coleman, Brown, & Rivkin, 1997; McMaster, Fuchs, & Fuchs, 2006;
Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003).

These findings suggest that acting as a peer model may also have
beneficial effects on learning. However, peer tutors’ learning gains
may stem from other factors than an explanation study intention
and actual explaining; peer tutors’ learning may also be affected by
the interaction with the tutee, who may ask questions that stimu-
late the peer tutor’s reflective knowledge-building in the process
of formulating an answer to those questions (Graesser, Person, &
Maglino, 1995; Roscoe & Chi, 2007). Peer models in video
modeling examples, on the other hand, are explaining to fictitious
peers who are not physically present.

To the best of our knowledge, the only study that has investi-
gated the effects of acting as a peer model by a) preparing and
studying learning materials and b) explaining what was learned by
creating a video-based modeling example, was conducted by Spies
et al. (2012). As part of a collaborative learning course, secondary
education students were asked to create a 5 min long video. The
authors concluded, based on students’ self-reports, that the video
creation process fostered both motivation and learning. However,
because of the lack of experimental control (e.g., no control group
for study intention and actually explaining) and reliance on self-
reports, no conclusions can be drawn from this study regarding the
effects of study intention and video creation on learning and
motivation.

1.2. The present study

The purpose of the current experiments was to investigate and
disentangle the effects of acting as a peer model on learning and
transfer and explore potential effects on self-efficacy and perceived
competence. A study intention of being able to explain a task to
others, and actually explaining that task during video creation, may
not only affect learning and transfer, but also how peer models view
and assess their own capabilities to perform that task. Again, it has
been shown that observing models may enhance self-efficacy
(Bandura, 1997; Schunk, 1987), but whether the process of acting

1 Note that other authors have used the term ‘instructional explanations’ in a
somewhat more restricted sense (e.g., Wittwer & Renkl, 2010).
as a peer model affects self-efficacy and perceived competence is not known. Exploring this question would be interesting, as self-efficacy beliefs have been hypothesized to underpin motivation, well-being, persistence, study behavior, and achievement (Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001). Perceived competence is closely related to self-efficacy, but whereas self-efficacy represents specific expectations and convictions for a certain situation, perceived competence refers to more general knowledge and perceptions that are likely more stable and enduring (Bong & Skaalvik, 2003; Hughes, Galbraith, & White, 2011; Klassen & Usher, 2010). Like self-efficacy, perceived competence has also been shown to have a significant bearing on students’ motivation and learning (Harter, 1990; Ma & Kishor, 1997).

In Experiment 1, secondary education students read a test on syllogistic reasoning (the content of which was new to them) with the study intention of being able to either successfully complete a test (condition A), or to successfully explain the learning materials to others (conditions B and C). Whereas students with an explanation study intention did not actually have to create a video in one condition (B), they did have to do so in the other condition (C). Students in condition C spent the last 5 min of the allocated study time on creating a video-based modeling example. Because a teaching experience will influence learning and motivation (Bargh & Schul, 1980; Benware & Deci, 1984; Renkl, 1995), the students in the video creation condition were not informed beforehand that they would actually be creating the video. Experiment 2 replicated Experiment 1 with university students, but the instructions were slightly adapted to rule out potential alternative explanations for the findings of Experiment 1.

We hypothesized that compared to studying for a test (condition A), studying with the intention of being able to successfully explain learning materials to others (i.e., without actually doing so; condition B) would facilitate a more active study approach that would benefit learning (Hypothesis 1a) and transfer (Hypothesis 1b), because of the aforementioned comprehension monitoring and self-explanation processes that the explanation intention might elicit, which could enhance understanding. Enhanced understanding is particularly beneficial for transfer performance (e.g., Van Gog et al., 2004).

Secondly, we hypothesized that actually creating a video-based modeling example (condition C) would further benefit learning (Hypothesis 2a) and transfer (Hypothesis 2b) above and beyond effects of study intention, in line with findings that show that providing explanations during tutoring (Cohen et al., 1982) or during small group discussions (Van Blankenstein et al., 2011) can aid learning. In sum, we predict the following pattern of results: condition $C > B > A$.

Mental effort in answering test questions will be analyzed to get more insight into the quality of learning outcomes (Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). Mental effort invested in the test provides an additional and more subtle indicator of the quality of cognitive schemas acquired under the different instructional conditions compared to performance measures alone (Van Gog & Paas, 2008). That is, when equal/higher performance is reached with lower/equal effort investment, cognitive schemas are more efficient (Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). We expect no differences in effort investment among conditions prior to learning, that is, in completing the pretest (Hypothesis 3a). We do, however, expect that the hypothesized higher performance on the posttest in the explanation conditions (see Hypothesis 1 and 2) will be reached with equal or lower effort investment (Hypothesis 3b).

Finally, potential effects of study intention and video-creation on self-efficacy (Question 4a) and perceived competence (Question 4b) were explored. On the one hand, when students in the explanation intention conditions feel they were successful in answering the questions they may have posed themselves while monitoring their learning and in explaining while making the video, the hypotheses regarding learning and transfer may also apply to self-efficacy and perceived competence (i.e., condition $C > B > A$). On the other hand, the opposite may also occur when students in the explanation conditions would become uncertain because of the questions they ask themselves. For instance, for difficult tasks with the added pressure of understanding the learning materials, anxiety has been shown to increase (Ross & DiVesta, 1976) and intrinsic motivation has been shown to decline (Renkl, 1995).

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants were 76 Dutch secondary education students (47 female) in the fourth year of pre-university education (age range 15–17 years), which is the highest level of secondary education in the Netherlands and has a six-year duration.

2.1.2. Design

The experiment consisted of four phases: pretest, study phase, immediate posttest, and delayed posttest. Participants were randomly assigned to one of three conditions: (1) Test Condition (study intention: be able to successfully complete a test; no video creation; $n = 27$), (2) Explanation Condition (study intention: be able to successfully explain to others; no video creation; $n = 25$), or (3) Explanation-Video Condition (study intention: be able to successfully explain to others; video creation; $n = 24$).

2.2. Materials

The tests and the text to be studied were paper-based.

2.2.1. Pretest

The pretest consisted of eight conditional syllogistic reasoning items. Each item (multiple choice) required participants to indicate whether the conclusion was ‘valid’ or ‘invalid’ (i.e., whether or not it logically followed from the two premises, which participants had to assume were true). For each of the four forms of syllogistic reasoning (i.e., 1) affirming the antecedent: valid; 2) affirming the consequent: invalid; 3) denying the antecedent: invalid; and 4) denying the consequent: valid), there were two test items: one with and one without belief-bias (i.e., the tendency to confirm based on prior beliefs of real world knowledge). The belief-bias makes it harder to judge whether a conclusion is valid or invalid (George, 1995; Newstead, Pollard, Evans, & Allen, 1992). To illustrate, the pretest item for affirming the antecedent without belief-bias was: “If a person works at the V&D [a Dutch department store], then that person is 16 years old or older. Elles works at the V&D. Conclusion: Elles is 16 years old or older.”, and the item for affirming the antecedent with belief-bias was: “If an event takes place in the 22nd century, then that event takes place in the future.” The maximum total score on the pretest was eight points (i.e., one point for each correct answer). For a description of the self-efficacy and perceived competence items used in the pretest, see Section 2.2.6.

2.2.2. Study text

The study text was 1930 words and six pages long. It described how one can judge for syllogism reasoning tasks whether or not a
conclusion logically follows from the two premises. The text was specifically created for the present experiment and throughout the text, the first and second premises were respectively addressed to as ‘the rule’ and ‘the observation’, after which the ‘the conclusion’ followed. There were two versions of the text, one for the Test Condition and one for both Explanation Conditions. These two versions only differed with regard to the study intention specific prompts placed at the end of each page (in the footer). These prompts were: “Can you apply the information from this page to complete a test?” (Test Condition) or: “Can you explain the information on this page to a fellow student? (Explanation Conditions).”

The text started with a general introduction to deductive reasoning and an overview of the four forms of syllogistic reasoning. The different forms of syllogistic reasoning were explained using the same recurrent example (affirming the antecedent: “If John sees a clown, then he is afraid. John sees a clown. Conclusion: John is afraid”; denying the antecedent: “If John sees a clown, then he is afraid. John does not see a clown. Conclusion: John is not afraid”; et cetera). How to judge whether a conclusion is valid or invalid was explained both in terms of the example and in abstract terms (e.g., for denying the antecedent: “If p, then q. not p therefore not q”), and for each form of syllogistic reasoning belief-bias, but explanations providing a specific example. The last page provided a table that summarized which forms of syllogistic reasoning were valid and invalid (see Appendix A), using another generic example (i.e., “If this is an apple, then it is fruit”).

2.2.3. Video creation
Participants in the Explanation-Video Condition were instructed on paper to 1) explain the four forms of syllogistic reasoning in front of a webcam, as if explaining them to someone without any knowledge on the subject, with the help of an adapted table taken from the study text (see Appendix B), and 2) explain what errors people commonly make when judging whether a conclusion is valid or invalid, and why people make this mistake (this refers to the belief-bias, but belief-bias was not explicitly mentioned in the instructions).

2.2.4. Posttests
The immediate and delayed posttests consisted of eight conditional syllogistic reasoning items (two for each form; one with and one without belief-bias) to assess learning outcomes, that is, retention of information from the experimental text about syllogisms. These test items were structurally equivalent but different in surface features compared to the pretest and again required participants to decide whether a conclusion was valid or invalid, but now they were also asked to explain their answer.

To assess transfer ([i.e., applying what had been learned from the text to new tasks], two Wason selection tasks were used (task 1: concrete context; task 2: abstract). The Wason selection tasks (Wason, 1966) ask people how they can test the validity of a rule such as “If a card has Y on one side, then it has 2 on the other side” by turning two cards from a set of four (e.g., showing X, Y, 2, 7). This task can be solved correctly if one understands the validity of the different reasoning forms as taught in the syllogism tasks and knows how to apply them. That is, to reach the correct solution, one needs to affirm the antecedent (turning Y) and deny the consequent (turning 7), but most people tend to affirm the antecedent (turning Y) as well as the consequent (turning 2). The Wason selection task items were introduced by stating that these were new tasks, but that what had been learned from the text could be used to successfully complete them. Again, participants were asked to select the right answer as well as to explain their answer.

Two parallel versions (A and B) of the posttest were created. Both test versions were structurally equivalent but different in surface features. On both posttests, the maximum score on the syllogistic reasoning items (i.e., learning outcomes) was 56 points. Each no belief-bias item was worth six points (one point for the correct choice on the multiple-choice question and five points for the explanation: one point for correctly recalling the form of syllogistic reasoning, one point for explaining correctly in abstract terms of p and q, two points for explaining correctly in concrete terms, and one point for correctly concluding in the explanation whether a conclusion was valid or invalid) and each belief-bias item was worth eight points (identical scoring to the no belief bias items, but with two points extra for correctly explaining the belief-bias).

Note that scoring of the explanations was not dependent on correctness of the initial answer, so some points could still be gained for correct elements in the explanation even when the answer regarding validity of the conclusion was incorrect. Participants could earn 18 points in total for the Wason selection tasks (i.e., nine per question: one point for selecting the correct answer, and two points per correct explanation for each of the four forms of syllogistic reasoning as applied to the rule in the Wason selection task).

Two raters independently scored 10% of the tests. The intra-class correlation coefficient was 0.949. Because of the high inter-rater reliability the remainder of the tests was scored by one rater.

2.2.5. Mental effort
Invested mental effort was measured after each test item on the pretest and both posttests using a subjective 9-point rating scale (Paas, 1992), which has a range from (1) very, very low effort to (9) very, very high effort. Average invested mental effort was computed separately for the syllogistic reasoning items and the Wason selection tasks.

2.2.6. Self-efficacy and perceived competence scale
Self-efficacy was measured using an adapted version of Bandura’s (2006) problem-solving self-efficacy scale, which asks participants to rate the degree of confidence (from 0% to 100%) in their ability to solve an incremental percentage of the total number of problems (from 10% to 100%, with increases of 10%). We altered this scale slightly by asking participants after the pretest to rate their degree of confidence (from 0% to 100%) in their ability to learn an in-depth explanation of the eight items (1 out of 8 to 8 out of 8, with increases of 1). Thus, participants firstly indicated their degree of confidence (0–100%) in their ability to learn an in-depth explanation of 1 out of 8 test items well, then the degree of confidence (0–100%) in their ability to learn an in-depth explanation of 2 out of 8 test items well, and so forth till 8 out of 8 test items. The wording of the self-efficacy measurements prior to each posttest was adjusted by asking participants to rate their confidence (from 0% to 100%) in their ability to answer an incremental amount of questions on a test (1 out of 8 to 8 out of 8, with increases of 1).

Perceived competence was measured using an adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996), which is a four item questionnaire that asks people to rate on a scale of 7 (not at all true) to 1 (very true): “I feel confident in my ability to learn this material”, “I am capable of learning the material in this course”, “I am able to achieve my goals in this course”, and “I feel able to meet the challenge of performing well in this course”. We altered this scale slightly by asking participants to rate the perceived competence in their capability to learn an in-depth explanation of the eight items (after the pretest) and in answering eight questions on a test (prior to each posttest). Both self-efficacy and perceived competence ratings were averaged for each measurement (i.e., after the pretest and before the immediate and delayed posttest).
2.3. Procedure

The study was run in two sessions. The first session (ca. 50 min) took place in the university lab. In a room with an entire class of students present, students were randomly allocated to a condition, after which the class was split-up in small groups of max. 10 participants according to condition; each group was supervised by an experiment leader. Participants in the Explanation-Video Condition were seated in sound-proof cubicles with the doors open so that they could hear and see the experiment leader; the doors were only closed once they started the actual video creation so that they would not hear each other. Participants in the other two conditions were either tested in sound-proof cubicles or a larger room with several seats between them. Participants received an envelope with instructions when creating the video, leaving 74 participants in total.

Participants according to condition; each group was supervised by an experiment leader. Participants in the Explanation-Video Condition received version B. The learning phase was over, participants were asked to take out the first booklet, containing the pretest and the first self-efficacy and perceived competence measures. They were given 12 min to complete the pretest. When time was up, the experiment leader asked participants to return the first booklet to the envelope and take out the second booklet, which contained the study test. The experimenter repeated the study intention prompt and indicated the amount of time participants would get to study the test. The Test and Explanation Condition received 17 min study time, while participants in the Explanation-Video Condition received 12 min study time. All participants were instructed several times to study for the full length of available study time. When time was up, participants were instructed to place booklet 2 back into the envelope. Participants in the Explanation-Video Condition were then told to turn on the webcam, and to create the video using the instructions that were handed out by the experiment leader. Note that participants in the Explanation-Video Condition were not aware prior to this point that they would be asked to create a video. Participants could see themselves on the computer screen during video creation. After the learning phase was over, participants were instructed that they would have 25 min for booklet 3. This booklet contained the self-efficacy and perceived competence scales, followed by the posttest. Half of the participants in each condition received version A as the immediate posttest while the other half received version B.

The second session took place 4 days later at participants’ schools, and lasted 25 min during which the delayed posttest was completed. A fourth booklet was handed out that again started with the envelope and take out the second booklet, which contained the study text. The experimenter repeated the study intention prompt, and indicated the amount of time participants would get to study for the full length of available study time. When time was up, participants were asked to take out the first booklet, containing the pretest and the first self-efficacy and perceived competence measures. They were given 12 min to complete the pretest. When time was up, the experiment leader asked participants to return the first booklet to the envelope and take out the second booklet, which contained the study test. The experimenter repeated the study intention prompt and indicated the amount of time participants would get to study the test. The Test and Explanation Condition received 17 min study time, while participants in the Explanation-Video Condition received 12 min study time. All participants were instructed several times to study for the full length of available study time. When time was up, participants were instructed to place booklet 2 back into the envelope. Participants in the Explanation-Video Condition were then told to turn on the webcam, and to create the video using the instructions that were handed out by the experiment leader. Note that participants in the Explanation-Video Condition were not aware prior to this point that they would be asked to create a video. Participants could see themselves on the computer screen during video creation. After the learning phase was over, participants were instructed that they would have 25 min for booklet 3. This booklet contained the self-efficacy and perceived competence scales, followed by the posttest. Half of the participants in each condition received version A as the immediate posttest while the other half received version B (and vice versa).

3. Results

Two participants from the Explanation-Video Condition were removed from all analyses because of non-compliance with the instructions when creating the video, leaving 74 participants in total.

The test scores on the learning (i.e., syllogism) and transfer (i.e., Wason selection) tasks in Experiment 1 are provided in Table 1, both in terms of score on the ‘multiple choice’ part of the questions (mc) and in terms of a ‘total’ (tot) score, which consists of the multiple choice score plus the score on the open-ended part of the question (explain your answer) taken together.

Five participants were absent at the delayed test (two from the Test Condition, one from the Explanation Condition, and two from the Explanation-Video Condition). For the delayed posttest scores of the five absent participants, a missing value analysis using the expectation maximization (EM) method was performed in SPSS. For the ANOVAs, Cohen’s $f$ is reported as a measure of effect size, with values of 0.10, 0.25, and 0.40 representing a small, medium, and large effect size, respectively. For the post-hoc tests, Cohen’s $d$ is reported with values of 0.20, 0.50, and 0.80 representing a small, medium, and large effect size, respectively (Cohen, 1988).

As expected, there were no differences among conditions in pretest performance, $F(2, 70) = 0.33, p = .720, f = 0.097$. To test our hypotheses regarding the effects on learning (Hypothesis 1a and 2a), a repeated measures ANOVA on the total performance scores on the syllogism tasks, with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor showed a main effect of Test Moment, $F(1, 70) = 4.67, p = .034, f = 0.256$. This indicated that participants performed somewhat better on the Immediate ($M = 22.18, SD = 6.66$) than on the Delayed Posttest ($M = 20.74, SD = 5.99$). There was also a significant effect of Instruction Condition, $F(2, 70) = 3.17, p = .048, f = 0.301$. Follow-up Bonferroni corrected post-hoc tests showed that, although the Explanation-Video Condition ($M = 23.94, SE = 1.19$) scored higher on average than the Explanation Condition ($M = 20.06, SE = 1.09$), this difference was not statistically significant, $p = .057, d = 0.710$. There were no differences between the Explanation-Video Condition and Test Condition ($M = 20.82, SE = 1.05$), $p = .159, d = 0.572$, nor between the Explanation Condition and Test Condition, $p = 1.00, d = 0.138$. Furthermore, there was no interaction between Test Moment and Instruction Condition, $F(2, 70) = 0.58, p = .561, f = 0.125$.

To test our hypotheses regarding the effects on transfer (Hypothesis 1b and 2b), a similar repeated measures ANOVA on the total performance scores on the Wason selection tasks showed no main effect of Test Moment (Immediate and Delayed Posttest), $F(1, 70) = 0.46, p = .499, f = 0.081$. There was a main effect of Instruction Condition, $F(2, 70) = 4.45, p = .015, f = 0.357$. Bonferroni corrected post-hoc tests on participants’ score over the Immediate and Delayed Posttest showed that performance in the Explanation-Video Condition ($M = 6.01, SE = 0.68$) was significantly better than in the Test Condition ($M = 3.55, SE = 0.60$), $p = .025, d = 0.791$, and the Explanation Condition ($M = 3.68, SE = 0.62$), $p = .041, d = 0.749$. No difference was found between the Test Condition and Explanation Condition, $p = 1.00, d = 0.042$, nor an interaction effect of Instruction Condition.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Learning</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test (mc)</td>
<td>Explanation (mc)</td>
</tr>
<tr>
<td>Pretest</td>
<td>5.00 (0.88)</td>
<td>4.80 (0.96)</td>
</tr>
<tr>
<td>Immediate</td>
<td>7.07 (1.36)</td>
<td>6.88 (1.27)</td>
</tr>
<tr>
<td>Immediate Posttest</td>
<td>21.07 (7.72)</td>
<td>21.17 (4.87)</td>
</tr>
<tr>
<td>Delayed</td>
<td>7.32 (1.17)</td>
<td>7.00 (1.19)</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>20.56 (6.33)</td>
<td>18.96 (5.76)</td>
</tr>
<tr>
<td>Pretest (tot)</td>
<td>2.74 (0.92)</td>
<td>2.87 (0.91)</td>
</tr>
<tr>
<td>Pretest (effort)</td>
<td>2.91 (1.35)</td>
<td>3.36 (1.07)</td>
</tr>
<tr>
<td>Delayed Posttest (effort)</td>
<td>2.43 (1.14)</td>
<td>3.21 (1.28)</td>
</tr>
</tbody>
</table>

### Table 1

Mean (SD) of learning and transfer multiple choice (mc) and total (tot) test scores and mental effort (effort) per condition in Experiment 1.
between Test Moment and Instruction Condition, $F(2, 70) = 0.57, p = .567, f = 0.127$.

### 3.2. Mental effort

Mental effort data are shown in Table 1. Six participants had one missing value on the syllogisms tasks (i.e., learning) and eight participants had a missing value on the Wason-selection tasks (i.e., transfer), which were replaced with the series mean. One participant with more than two missing values on the syllogism tasks and two participants with more than one missing value on the Wason-selection tasks were removed from the analyses. As expected (Hypothesis 3a), an ANOVA showed no significant differences among conditions in the mean invested mental effort during the pretest, $F(2, 70) = .86, p = .226, f = 0.015$. On the posttest (Hypothesis 3b), a repeated measures ANOVA on the average mental effort invested on the syllogism tasks with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor and Instruction Condition as between-subjects factor showed a main effect of Test Moment, $F(1, 63) = 13.63, p < .001, f = .456$, with participants investing more mental effort during the Immediate Posttest ($M = 3.12, SD = 1.25$) than during the Delayed Posttest ($M = 2.80, SD = 1.30$). There was no effect of Instruction Condition, $F(2, 63) = 1.54, p = .223, f = .221$, nor an interaction effect, $F(2, 63) = 1.31, p = .278, f = .185$. On the Wason selection tasks, there was also a main effect of Test Moment, $F(1, 62) = 13.67, p < .001, f = .468$. Again, participants invested more mental effort during the Immediate Posttest ($M = 3.65, SD = 1.45$) than during the Delayed Posttest ($M = 3.00, SD = 1.39$). There was no main effect of Instruction Condition, $F(2, 62) = 1.54, p = .223, f = .223$, nor an interaction effect, $F(2, 62) = .26, p = .773, f = .083$.

### 3.3. Self-efficacy and perceived competence

The self-efficacy and perceived competence data are provided in Table 2. Because the questionnaires that measured self-efficacy and perceived competence prior to the immediate and delayed posttests differed in phrasing compared to the pretest (see Section 2.2.6.), we analyzed them separately to answer Questions 4a and 4b. For the self-efficacy measure conducted after the pretest, seven participants had to be excluded because of missing values (leaving $n = 25$ in the Test, $n = 23$ in the Explanation, and $n = 19$ in the Explanation-Video Condition). An ANOVA showed no significant differences among conditions on the self-efficacy measurement after the pretest, $F(1, 64) = 2.13, p = .127, f = .258$. A repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor on the self-efficacy ratings provided before the Immediate and Delayed Posttest showed a main effect of Test Moment, $F(1, 63) = 18.39, p < .001, f = .528$, indicating that self-efficacy was significantly lower at the Delayed Posttest ($M = 75.30, SD = 15.81$) than at the Immediate Posttest ($M = 80.60, SD = 14.13$). There was no main effect of Instruction Condition, $F(2, 63) = 1.80, p = .173, f = .239$, nor an interaction effect, $F(2, 63) = 1.44, p = .245, f = .188$.

The adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996) used in the present study showed high internal consistency on the Pretest (Cronbach’s $\alpha = .925$), Immediate Posttest (Cronbach’s $\alpha = .920$), and Delayed Posttest (Cronbach’s $\alpha = .947$). Perceived competence after the pretest differed significantly among conditions, $F(1, 70) = 3.90, p = .025, f = 0.334$. Because the experimental procedure for both Explanation Conditions prior to the learning phase was still identical, a contrast test was conducted to investigate whether the Test Condition differed on perceived competence from both Explanation Conditions taken together after the pretest (i.e., after only one study intention prompt). Results showed a significant difference, $t(71) = 2.15, p = .035, d = 0.509$, with the Test Condition showing higher perceived competence ($M = 6.04, SE = 0.16$) than the Explanation Conditions ($M = 5.39, SE = 0.17$). A repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor on perceived competence ratings provided before the Immediate and Delayed Posttest, showed a main effect of Test Moment, $F(1, 64) = 18.99, p < .001, f = .530$, indicating a decrease in perceived competence from the Immediate ($M = 6.19, SD = 0.72$) to the Delayed Posttest ($M = 5.88, SD = 0.79$). Although the Test Condition seemed to have higher mean scores on perceived competence, there was no statistically significant effect of Instruction Condition, $F(2, 64) = 2.88, p = .063, f = .200$. There was no statistically significant interaction effect between Test Moment and Instruction Condition, $F(2, 64) = 1.76, p = .181, f = .206$.

### 4. Discussion

In contrast to our first hypothesis, studying with the intention of being able to successfully explain materials to others (i.e., Explanation Condition) was not more beneficial for learning (Hypothesis 1a) and transfer (Hypothesis 1b) than studying to be able to successfully complete a test (i.e., Test Condition). Regarding our second hypothesis, while actually explaining to non-present others by creating a video-based modeling example (i.e., Explanation-Video Condition) did not have an effect on learning (Hypothesis 2a), it did have a significant beneficial effect on transfer (Hypothesis 2b). This beneficial effect on transfer performance was reached with the same investment of mental effort on the posttest (Hypothesis 3b), indicating that the cognitive schemas acquired in the Explanation-Video Condition were not only more effective, but also more efficient (Van Gog & Paas, 2008).

The explorative analysis of potential effects that study intention and creating a video-based modeling example might have on self-efficacy and perceived competence (Questions 4a and 4b), showed no effects of instructional condition on self-efficacy. As for perceived competence, there was a significant difference between the Test Condition and both Explanation Conditions taken together after the pretest, that is, after just one study intention prompt. On the posttests, even though studying new materials with the intention of completing a test seemed to lead to a somewhat higher confidence in one’s capabilities, there were no statistically significant differences among conditions.

In sum, the results of Experiment 1 show that the process of explaining during video creation is particularly effective for enhancing transfer performance, which suggests that acting as a

---

**Table 2**

Mean (SD) of self-efficacy and perceived competence scores per condition in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Self-efficacy</th>
<th>Perceived Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Explanation</td>
</tr>
<tr>
<td>Pretest</td>
<td>80.67 (12.55)</td>
<td>73.94 (13.33)</td>
</tr>
<tr>
<td>Immediate Posttest</td>
<td>83.45 (11.02)</td>
<td>79.20 (15.82)</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>80.78 (13.62)</td>
<td>71.57 (16.04)</td>
</tr>
</tbody>
</table>
peer model for a video-based modeling example can be an effective educational activity in and of itself.

Based on Experiment 1, however, it is not entirely clear whether this positive effect was ‘simply’ caused by recalling information from long-term memory (i.e., retrieval practice) or specifically by explaining during video creation. That is, only participants in the Explanation-Video Condition had to recall information from memory, and they were given a table that they could use to explain the four types of syllogisms. Engaging in (cued) retrieval practice has been shown to positively affect learning outcomes (Karpicke & Blunt, 2011). Furthermore, the Explanation-Video Condition also instructed the participant to focus on what errors people commonly make when judging whether a conclusion is valid or invalid, and why people make this mistake (i.e., the belief-bias). Although the study text described the belief-bias repeatedly throughout the text, we cannot rule out the possibility that the emphasis on explaining this error has influenced the transfer results.

To address these issues, a second experiment was conducted. In Experiment 2 we investigated whether we could replicate the results from Experiment 1 with university students who were also novices on the topic of syllogistic reasoning, while ruling out retrieval practice and the focus on common errors as factors that could explain the positive findings of the process of acting as a peer model in Experiment 1.

5. Experiment 2

In Experiment 2, after reading the text for 12 min, participants in the Test and Explanation Condition also engaged in a short cued recall activity prior to continuing studying of the text. They were instructed to fill in the gaps of the table in Appendix B from memory (i.e., to apply the example ‘If this is an apple, then it is fruit’ to the four forms of syllogistic reasoning) to ensure that all three conditions engaged in retrieval practice after studying for 12 min. Furthermore, the Explanation-Video Condition no longer received the instruction to explain during video creation what errors people commonly make when judging whether a conclusion is valid or invalid, and why people make this mistake.

If the positive effects of the process of acting as a peer model for a video-based modeling example in Experiment 1 arose because of retrieval practice and/or focusing on the belief-bias during video creation, then no differences among the conditions should be found in Experiment 2. However, as in Experiment 1, we hypothesize that the key processes in the Explanation-Video Condition of a) studying with the intention of explaining to others later on and b) actually explaining during video creation, will have beneficial effects on learning outcomes, particularly on transfer. The reader is referred to Section 1.2 for a description of hypotheses and research questions.

5.1. Participants

Participants were 95 Dutch undergraduate students (age $M = 20.41, SE = 0.19$; 65 female; 90 of those participants studied Psychology in a Problem-Based Learning curriculum). Students who had studied or were studying philosophy or students with programming experience were not eligible for participation to ensure that the population consisted of novices with regard to the topic of syllogistic reasoning. Participants were rewarded for their participation with course credits or a monetary reward.

5.2. Design

The design was the same as in Experiment 1. Participants were randomly assigned to one of three conditions: (1) Test Condition ($n = 32$), (2) Explanation Condition ($n = 31$), or (3) Explanation-Video Condition ($n = 32$).

5.3. Materials and procedure

The same materials and procedure were used as in Experiment 1, with a few exceptions.

Firstly, after the text (booklet 2) that was studied for 12 min, a booklet 3 was handed out that presented participants in all conditions with the table taken from the study text, although it was now no longer filled out as it was in the study text (see Appendix B). In the Test Condition and Explanation Condition, participants were instructed to fill in the gaps in the table from memory. If they were finished, they were instructed to proceed to turn the page and study the rest of booklet 3, which presented the text again for restudy. They were given max. 5 min. in total for recall and restudy. To investigate whether retrieval from long-term memory was successful (i.e., how well students had filled in the gaps in the table), recall performance was scored. Participants could earn a total of 8 points, 2 points per correctly applied form of syllogistic reasoning. They received 0 points for a wrong answer, 1 point for a partially correct answer, and 2 points for a fully correct answer. In the Explanation-Video Condition, the table was preceded by the instruction to use it while explaining the four forms of syllogistic reasoning in front of a webcam, as if explaining them to someone without any knowledge on the subject. Participants were given maximally 5 min for creating the video.

Secondly, the study intention prompt during the general instruction at the start of the experiment was removed to ensure that the self-efficacy and perceived competence measures made after the pretest could act as control variables for the self-efficacy and perceived competence measures made prior to both posttests. Thirdly, all participants were tested in individual cubicles with maximally eight participants per session during both sessions. Finally, the delayed test took place 7 days later (instead of 4 days in Experiment 1).

Posttest performance was scored in the same manner as in Experiment 1. Two raters independently scored 10% of the posttests. The intra-class correlation coefficient was 0.898, and because of the high agreement the remainder of the tests was scored by one rater.

6. Results

One participant from the Explanation-Video Condition was removed from all analyses because of non-compliance with the instructions.

To assess the quality of recall (i.e., filling in the gaps in the table depicted in Appendix B) in the Test and Explanation Condition, an independent samples $t$-test was conducted. There was no statistically significant difference between the Test Condition ($M = 7.84, SD = 0.63$) and the Explanation Condition ($M = 7.74, SD = 0.86$), $t(61) = 0.54$, $p = .591$, $d = 0.136$. Both were highly successful (the maximum score that could be obtained was 8) in filling in the gaps in the table from memory.

6.1. Learning outcomes and transfer test performance

The multiple choice and total (i.e., mc + explanation) test scores on the learning (i.e., syllogism) and transfer (i.e., Wason selection) tasks in Experiment 2 are displayed in Table 3. Three participants from the Test Condition were absent at the delayed test. An EM missing value analysis in SPSS was performed to replace their missing values.

An ANOVA showed no significant differences among conditions in pretest performance, $F(2, 91) = 0.06$, $p = .940$, $f = 0.037$. A
repeated measures ANOVA on the total performance scores on the syllogism tasks (Hypothesis 1a and 2a), with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor showed a main effect of Test Moment, $F(1, 91) = 8.30, p = .005, f = .300$, indicating a decrease in learning performance from the Immediate ($M = 24.64, SD = 6.45$) to the Delayed Posttest ($M = 22.80, SD = 7.35$). There was also a significant main effect of Instruction Condition, $F(2, 91) = 9.99, p < .001, f = .469$. Bonferroni corrected post-hoc tests showed that the Test Condition ($M = 20.25, SE = 1.00$) was significantly outperformed by both the Explanation-Video Condition ($M = 26.49, SE = 1.02$), $p < .001, d = 1.100$, and the Explanation Condition ($M = 24.52, SE = 1.02$), $p = .011, d = 0.753$. There was no significant difference between the Explanation-Video Condition and Explanation Condition, $p = .526$, $d = 0.347$, nor was there an interaction effect between Test Moment and Instruction Condition, $F(2, 91) = 0.68, p = .507, f = .017$.

To test our hypotheses regarding the effects on transfer (Hypothesis 1b and 2b), a similar repeated measures ANOVA on the total performance scores on the Wason selection tasks showed no main effect of Test Moment (Immediate and Delayed Posttest), $F(1, 91) = 1.16, p = .283, f = .0110$. There was a main effect of Instruction Condition, $F(2, 91) = 4.04, p < .021, f = .298$. Bonferroni post-hoc tests showed no difference between the Explanation-Video Condition ($M = 6.70, SE = 0.48$) and Explanation Condition ($M = 6.39, SE = 0.48$), $p = 1.000, d = 0.117$, nor between the Explanation Condition and Test Condition ($M = 4.90, SE = 0.48$), $p = .093, d = 0.552$. However, as expected, performance in the Explanation-Video Condition was significantly higher than in the Test Condition, $p = .028, d = 0.700$. No interaction effect was found between Test Moment and Instruction Condition, $F(2, 91) = 2.24, p = .112, f = .221$.

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Learning</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Explanation</td>
</tr>
<tr>
<td>Pretest (mc)</td>
<td>5.47 (1.50)</td>
<td>5.52 (1.12)</td>
</tr>
<tr>
<td>Immediate Posttest (mc)</td>
<td>7.44 (1.19)</td>
<td>7.58 (0.85)</td>
</tr>
<tr>
<td>Immediate Posttest (tot)</td>
<td>21.69 (6.39)</td>
<td>25.15 (5.78)</td>
</tr>
<tr>
<td>Delayed Posttest (mc)</td>
<td>6.88 (1.76)</td>
<td>7.26 (1.26)</td>
</tr>
<tr>
<td>Delayed Posttest (tot)</td>
<td>18.82 (6.51)</td>
<td>23.90 (5.66)</td>
</tr>
<tr>
<td>Pretest (effort)</td>
<td>3.48 (1.09)</td>
<td>3.20 (1.15)</td>
</tr>
<tr>
<td>Immediate Posttest (effort)</td>
<td>2.87 (1.33)</td>
<td>2.57 (1.09)</td>
</tr>
<tr>
<td>Delayed Posttest (effort)</td>
<td>2.91 (1.39)</td>
<td>2.43 (1.12)</td>
</tr>
</tbody>
</table>

6.2. Mental effort

Mental effort data are shown in Table 3. For one participant, a missing value on the pretest was replaced with the series mean. On the posttests, five participants had one missing value on the syllogisms tasks (i.e., learning) and five participants had a missing value on the Wason selection tasks (i.e., transfer), all of which were replaced with the series mean. One participant with more than two missing values on the syllogism tasks was removed from the analysis on learning. As expected (Hypothesis 3a), an ANOVA showed no significant differences among conditions on the mean invested mental effort during the pretest, $F(2, 91) = 1.87, p = .159, f = 0.203$. For the average mental effort invested on the syllogism tasks, a repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor and Instruction Condition as between-subjects factor showed no main effect of Test Moment, $F(1, 87) = 0.01, p = .913, f < .001$. There was, however, a significant main effect of Instruction Condition, $F(2, 87) = 3.471, p = .035, f = 0.282$. Bonferroni corrected post-hoc tests showed that participants in the Explanation-Video Condition ($M = 3.37, SE = 0.24$) invested significantly more mental effort than participants in the Explanation Condition ($M = 2.50, SE = 0.23$), $p = .030$, $d = 0.675$. No differences were found between the Explanation-Video Condition and the Test Condition ($M = 2.89, SE = 0.24$), $p = .469$, $d = 0.373$, nor between the Explanation Condition and Test Condition, $p = .740$, $d = 0.301$. No interaction effect was found between Test Moment and Instruction, $F(2, 87) = 0.68, p = .510, f = .125$.

For the average mental effort invested on the Wason selection tasks, a repeated measures ANOVA showed a main effect of Test Moment, $F(1, 88) = 19.49, p < .001, f = 0.463$. With invested mental effort being significantly lower on the Delayed Posttest ($M = 3.88, SD = 1.92$) than on the Immediate Posttest ($M = 4.65, SD = 1.89$). There was no main effect of Instruction Condition, $F(2, 88) = 1.14, p = .325, f = 0.161$, nor an interaction effect, $F(2, 88) = 1.44, p = .242, f = .164$.

6.3. Self-efficacy and perceived competence

The self-efficacy and perceived competence data are provided in Table 4. Note that in the present experiment the self-efficacy and perceived competence estimates made after the pretest acted as control variables as a result of the removal of the study prompt prior to the pretest. Because the questionnaires that measured self-efficacy and perceived competence prior to the immediate and delayed posttests differed in phrasing compared to the pretest (see Section 2.2.6.), we analyzed them separately to answer Questions 4a and 4b. For self-efficacy, an ANOVA showed no significant differences among conditions on the measurement after the pretest, $F(2, 91) = 0.40, p = .669, f = 0.094$. A repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor on the self-efficacy ratings provided before the Immediate and Delayed Posttest showed a main effect of Test Moment, $F(1, 88) = 6.00, p = .016, f = 0.258$, indicating that self-efficacy significantly decreased from the Immediate Posttest ($M = 81.65, SD = 14.64$) to the Delayed Posttest ($M = 78.73, SD = 16.52$). There was no main effect of Instruction Condition, $F(2, 88) = 1.75, p = .180, f = 0.199$, nor an interaction effect, $F(2, 88) = 1.10, p = .339, f = 0.153$.

The adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996) again showed high internal consistency on the Pretest (Cronbach’s $\alpha = .842$), Immediate Posttest (Cronbach’s $\alpha = .945$), and Delayed Posttest (Cronbach’s $\alpha = .934$). Perceived competence after the pretest did not differ significantly among conditions, $F(2, 91) = 2.07, p = .133, f = .213$. A repeated measures ANOVA showed a significant effect of Test Moment (Immediate Posttest and Delayed Posttest), $F(1, 88) = 4.30, p = .041, f = 0.220$. Participants’ perceived competence was significantly lower on the Delayed Posttest ($M = 6.00, SD = 0.89$) than on the Immediate Posttest ($M = 6.13, SD = 0.82$). There was also a significant main effect of Instruction Condition, $F(2, 88) = 5.23, p = .007, f = 0.345$. Bonferroni corrected post-hoc tests showed that the
Explanation Condition ($M = 6.36, SE = 0.14$) scored higher on perceived competence than the Explanation-Video Condition ($M = 5.73, SE = 0.14$), $p = .005$, $d = 0.818$. No difference was found between the Explanation Condition and Test Condition ($M = 6.10, SE = 0.14$), $p = .575$, $d = 0.340$, nor between the Explanation-Video Condition and the Test Condition, $p = .203$, $d = 0.477$. There was no interaction between Test Moment and Instruction Condition, $F(2, 88) = 0.25, p = .782, f = 0.073$.

7. Discussion

Congruent with our hypothesis, studying with the intention of being able to successfully explain materials to others without actually doing so (i.e., Explanation Condition) was more beneficial for learning (Hypothesis 1a) than studying to successfully complete a test (i.e., Test Condition). However, actually explaining to non-present others during video creation (i.e., Explanation-Video Condition) was not more effective for learning than studying to explain to others without actually doing so (Hypothesis 2a). Interestingly, actually having explained to others did result in higher mental effort investment on the posttest than only having had the intention to explain to others or the intention to study for a test. This suggests that with regard to learning, the Explanation Condition had acquired more efficient cognitive schemata than the Explanation-Video Condition (Van Gog & Paas, 2008).

This second experiment again showed positive effects on transfer of the processes involved in acting as a peer model relative to studying with the intention of successfully completing a test. No effect of study intention on transfer was found (Hypothesis 1b), but in line with our hypothesis, actual video creation significantly fostered transfer performance compared to studying for a test (Hypothesis 2b). This higher transfer performance was reached with equal investment of mental effort on the posttest as in the other conditions (Hypothesis 3b), suggesting that the cognitive schemas acquired in the Explanation-Video Condition were not only more effective, but also more efficient compared to the Test Condition (Van Gog & Paas, 2008).

Our exploration of self-efficacy (Hypothesis 4a) showed no differences among the conditions. However, studying with the intention of being able to successfully explain materials to others without actually doing so led to higher perceived competence than studying to explain to others followed by video creation (Hypothesis 4b).

In sum, Experiment 2 showed a positive effect of studying with the intention of explaining to others on learning. Moreover, Experiment 2 provided additional evidence for the notion that engaging in the two processes involved in acting as a peer model for a video-based modeling example (i.e., first studying with the intention of explaining later on and then actually explaining on webcam) are more effective for fostering transfer than studying with the intention of completing a test, while ruling out retrieval practice and the focus on common errors as factors for these positive effects.

8. General discussion

The present experiments investigated the effects of study intention and creating a video-based modeling example on learning, transfer, and mental effort, and explored the effects on self-efficacy, and perceived competence. Experiment 1 showed no differences between studying to be able to successfully complete a test (i.e., Test Condition) and studying with the intention of being able to successfully explain materials to others without actually doing so (i.e., Explanation Condition) on learning (Hypothesis 1a) and transfer (Hypothesis 1b). Experiment 2, however, did show an effect of study intention on learning, but not for transfer. For university students, a study intention of explaining to others later on was more beneficial for learning than studying for a test. There are two potential explanations for this finding. First, the fact that the majority of participants in Experiment 2 were enrolled in a PBL curriculum means that they had a substantial amount of experience with explaining studied materials to others (a post-discussion of the literature that was studied at home is the last step of the PBL model used in this curriculum; Loyens, Kirschner, & Paas, 2012). Thus, these findings might suggest that the secondary education students may not have fully adopted the unfamiliar explanation study intention, while the university students for whom this was a more familiar approach, did adopt it. Second, in Experiment 2, participants in the Test and Explanation Condition also had to engage in memory retrieval, which was not the case in Experiment 1. Possibly, it was the explanation study intention combined with retrieval practice that fostered learning.

In line with our second hypothesis, actually explaining to non-present others by creating a video-based modeling example (i.e., Explanation-Video Condition) had an effect on learning (Hypothesis 2a) and on transfer (Hypothesis 2b). In Experiment 1, the process of video creation fostered transfer (Hypothesis 2b) relative to studying to successfully complete a test. In Experiment 2, video creation was more effective for both learning and transfer compared to studying for a test. Moreover, whereas the Explanation-Video Condition significantly outperformed the Explanation Condition on transfer performance in Experiment 1, this was not the case in Experiment 2. Similar to the findings with regard to learning outcomes, the better transfer performance of the Explanation Condition in Experiment 2 compared to Experiment 1, could be a result of participants’ experience with a study intention of explaining to others later on and/or of engaging in memory retrieval before restudying. Nevertheless, the Explanation Condition did not perform significantly better on transfer than the Test Condition ($d = 0.552$), only the Explanation-Video Condition did. These results show that actually providing explanations to others not only has positive effects on learning in interactive situations such as tutoring (Cohen et al., 1982) and small group discussions (Van Blankenstein et al., 2011), but also in the non-interactive situation of explaining while creating a video-based modeling example. Moreover, the results showed that explaining mainly affects transfer, which was not measured in earlier studies.

In sum, two experiments showed that explaining to non-present others during video creation resulted in better transfer performance relative to studying for a test. To be able to solve and explain the Wason selection tasks correctly (i.e., transfer), students had to understand the validity of the different reasoning forms as taught in the syllogism tasks and had to know how to apply this.
information to a new task. Apparently, explaining during video creation helped participants in both experiments to see the structural analogy between the two types of tasks and realize how the syllogistic reasoning rules could be successfully applied to the Wason selection tasks (for a discussion of the importance of perceiving analogies for transfer, see Gick & Holyoak, 1983; Needham & Begg, 1991; Renkl, 2011). But what was it about explaining aloud during video creation that helped participants see this analogy? The results of Experiment 2 show that it is not simply an effect of recalling information from memory, nor a result of focusing on the belief-bias during video creation. It is likely that actively explaining to non-present others helped participants to make connections between different elements of information and to process information on a deeper level, resulting in improved understanding, which has been suggested to be particularly beneficial for transfer performance (Van Gog et al., 2004). Furthermore, speaking aloud while explaining may also have been a contributing beneficial factor. Research on the production effect has shown that producing a word aloud can increase its distinctiveness and therefore improve explicit memory compared to studying in silence (e.g., MacLeod, Gopie, Hounihan, Neary, & Oszubko, 2010). Thus far, however, it is unclear whether the beneficial effect can be generalized from learning vocabulary to learning higher-order skills.

The results on invested mental effort show that participants who studied with the intention of being able to successfully explain materials to others without actually doing so invested significantly less mental effort in completing the posttest items that measured learning than participants who did explain to others during video creation, while both conditions performed equally on those items. This indicates that the Explanation Condition acquired more efficient cognitive schemas (Hypothesis 3b; see Van Gog & Paas, 2008), a surprising finding that is hard to explain. Congruent with our expectations, however, both Experiment 1 and 2 showed that the beneficial effects of explaining during video creation on transfer performance were reached with the same investment of mental effort on the transfer items. This indicates that, in line with our hypothesis, the acquired cognitive schemas of students who actually explained to others during video creation were not only more effective but also more efficient for transfer performance than those of students in the other conditions.

It would be interesting for future research to also measure the invested mental effort during the learning phase, because it could be hypothesized that actually explaining on video might require more mental effort as a learning strategy than studying for a test or studying with the intention to explain but not actually doing so. In other words, it might be an activity that imposes germane, or effective, cognitive load (Sweller et al., 1998). Measuring invested effort during or shortly after the learning phase would ideally have to be in addition to measuring effort invested in the test though, because mental effort ratings during the learning phase can be difficult to interpret on their own. Not only are such ratings directly affected by the instructional conditions, but levels of mental effort invested during learning cannot tell us whether this effort is evoked by processes relevant (i.e., germane load) or irrelevant (i.e., extraneous) for learning without looking at the quality of the cognitive schemas that were acquired (Van Gog & Paas, 2008).

Our explorative analysis showed no effects of study intention and creating a video-based modeling example on self-efficacy (Question 4a). Self-efficacy did however significantly decrease from immediate to delayed posttest in both experiments. Effects on perceived competence (Question 4b) were found in both experiments. The results of Experiment 1 showed that, for secondary education students, receiving a prompt to study new materials with the intention of successfully completing a test led to more confidence in their own capabilities after completing a pretest (i.e., before the learning phase). After the learning phase, the differences among conditions were not statistically significant. For the university students in Experiment 2—in which the study intention prompt that was given during the general instruction at the start of the experiment was removed—the results suggest that the process of video creation lowered confidence in one’s own capabilities in comparison to studying with the intention of explaining later on without actually explaining, but not compared to studying with the intention to complete a test. A possible explanation for this finding might be that a more active approach to learning, such as explaining during video creation, can make novice learners more aware of the difficulty of the materials (see also findings that teaching expectancy may lead to increased anxiety, Ross & DiVesta, 1976, and decreased intrinsic motivation, Renkl, 1995), while they may not realize that these difficulties may be conducive to performance on a subsequent test. Another possible explanation is the degree to which the activities in the learning phase were new to the students; the majority of university students had experience both with test and explanation study intentions (due to their PBL curriculum), but not with video creation. However, these explanations are tentative at best, because we found no such effect in Experiment 1 and only found the effect compared to the Explanation Condition, not compared to the Test Condition in Experiment 2. Considering the question what differences between conditions in perceived competence mean exactly, it would be interesting in future research to investigate whether students’ beliefs about their abilities are accurate considering their level of test performance.

Our findings are also of interest for educational practice. The use of video-based instruction in general is increasing in formal education, with educators even arguing for “flipping the classroom”, that is, having learners study video lessons at home, and using teaching time in school for practice and teacher support (Bergmann & Sams, 2012). Moreover, the use of video-based modeling examples is also increasing in both formal and informal learning (see e.g., Khanacademy.org). Asking students to create video-based modeling examples for other students does not only provide a more cost-efficient alternative to using teachers as models, but our results show it might also be an effective learning activity in and of itself. However, before educators can start relying on modeling examples created by peer students, research should first verify whether the videos that students create when acting as a peer model constitute effective educational materials. Whereas previous research does indeed suggest that video modeling examples created by peers are effective for students to learn from (e.g., Groenendijk et al., 2013a, 2013b), the examples in those studies focused on different tasks and were created under very different circumstances from those in the current study. Before videos of the kind created in this study can be used in educational practice, students may have to be allowed to design, edit, or re-do the video as well.

Future research could investigate whether the positive effects of the processes involved in acting as a peer model on learning and transfer only apply to students learning reasoning tasks, or if they can be generalized to other learning materials and task types, such as problem solving tasks. Possibly, the learning effects might even increase when students would be given more time to think about how to present the material to others (allowing them to design and edit the video itself); research in other domains has shown that learning by designing hypermedia (Lehrer & Romberg, 1996; Penner, Lehrer, & Schauble, 1998) and by designing “slow animations” (Hoban, Loughran, & Nielsen, 2011) can have positive effects on learning outcomes. Future research could also vary the degree of support that peer models receive during video creation. For example, peer models could be instructed to follow a predetermined script, which has been shown to scaffold knowledge acquisition in the interactive domain of computer-supported collaborative learning (Fischer, 2013).
Moreover, varying the degree of support by using scripts could help answer the question whether the process of explaining is already sufficient to produce the performance benefits. Another question that future research could explore is how learners’ conceptions of the posttest affect their study behavior and consequently the quality of acquired schemata. Possibly, if learners would be explicitly told that successful performance on the test relies on deep processing, they might study in a different manner and such a test study intention condition may become more similar to an explanation study intention. In addition, as mentioned above, future research should further investigate effects on self-efficacy and perceived competence, but should also try to address the question of whether students’ beliefs about their abilities are accurate considering their level of test performance. Finally, it remains an open question whether the presence of others would affect the quality of explanations and the confidence novice learners have in their own abilities. Whereas a peer tutor directly explains to the tutee, and consequently is fully aware of who the recipient of the explanation is, a peer model for a video-based modeling example may not be aware who the recipient is because the explanation is given indirectly. Further research on these issues would seem potentially fruitful for both educational theory and practice.

Acknowledgments

This research was funded by Kennisnet. The authors would like to thank Mee van Krimpen, Rola Hulsbergen, Marry Schuddebeurs, and Janneke Louwerse for facilitating this study, and Eveline Stoker, dien van Eersel, Jim Maarseveen, Jay Jay Mast, Lisette Wijnia, and Janneke Louwerse for helping with the data collection. We would also like to thank the three anonymous reviewers whose comments helped us improve the manuscript.

Appendix A. Summary table of syllogistic reasoning using a generic example presented in the study text

<table>
<thead>
<tr>
<th>If this is an apple, then it is fruit</th>
<th>Affirm</th>
<th>Deny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antecedent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If it is an apple</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Then it is fruit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consequent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If it is fruit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Then it is not a fruit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix B. Adapted summary table of syllogistic reasoning with the observation and conclusion removed, used in video-explanation condition in Experiment 1, and in all conditions in Experiment 2

<table>
<thead>
<tr>
<th>If this is an apple, then it is fruit</th>
<th>Affirm</th>
<th>Deny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antecedent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consequent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.


