Observe and explore: empirical studies about learning in creative writing and the visual arts

Groenendijk, T.

Publication date
2012

Chapter 3

THE EFFECTS OF OBSERVATIONAL LEARNING ON STUDENTS' PERFORMANCE, PROCESSES, AND MOTIVATION IN TWO CREATIVE DOMAINS

Previous research has shown that observation can be effective to improve learning in various domains, e.g. argumentative writing and mathematics. The question in this paper is whether observational learning can also be beneficial when students are learning to perform creative tasks in visual and verbal arts. We hypothesized that observation has a positive effect on performance, process and attitudes (intrinsic motivation, task value and self-efficacy). We expected similarity in competence between the model and the observer to influence the effectiveness of observation. A total of 131 Dutch students (10th grade, 15 years old) participated in two experiments (one for visual and one for verbal arts). Participants were randomly assigned to one of three conditions: two observational learning conditions and a control condition (learning by practising). The observational learning conditions differed in instructional focus (on the weaker or the more competent model of a pair to be observed). We found positive effects of observation on creative products, creative processes and task value in the visual domain. In the verbal domain, observation seemed to affect the creative process, but not the other variables. The model similarity hypothesis was not confirmed. Results suggest that observation may foster learning in creative domains, especially the visual arts.

1. INTRODUCTION

Before the introduction of formal education apprenticeship was the most common means of learning (Collins, Brown, & Newman, 1989). Apprenticeship includes modelling: an expert demonstrates his work process to an observing apprentice. Observational learning, as examined in the present study, is also triggered by a form of modelling; students learn by watching, interpreting and evaluating peers carrying out a task. In formal education, observational learning proved to be an effective learning activity in various domains, such as mathematics (e.g. Schunk & Hanson, 1985), reading (Couzijn & Rijlaarsdam, 2004), argumentative writing (Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Couzijn, 1999; Raedts, Rijlaarsdam, Van

This chapter is a slightly adapted version of a paper published as: Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2011). The effect of observational learning on students' performance, processes, and motivation in two creative domains. British Journal of Educational Psychology. *DOI: 10.1111/j.2044-8279.2011.02052.x*

Waes, & Daems, 2007; Rijlaarsdam, Braaksma, Couzijn, Janssen, Raedts, Van Steendam, Toorenaar, & Van den Bergh, 2008; Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010; Zimmerman & Kitsantas, 2002), learning to collaborate (Rummel & Spada, 2005) and learning argumentation skills (Schworm & Renkl, 2007).

It is generally believed that creative skills should be nurtured in the classroom. We expect secondary school students to display creative and independent thinking when working on projects, writing papers, or participating in classroom discussions. In arts education, teachers aim at stimulating students' creative processes. As far as we know, there are no experimental studies examining the effectiveness of interventions that aim at enhancing creative processes in secondary arts education. Implementing observational learning tasks may be an effective approach for stimulating students' creative processes and improving their creative products. In the following sections, we will describe the potential for learning from observation for creative tasks.

1.1 Observational learning

The rationale of observational learning is based on Bandura's (1986) social learning theory. It explains learning as a continuous interaction between cognitive, behavioural and environmental influences. For effective modelling, several conditions need to be fulfilled: students need to pay attention to relevant elements in the learning environment, to store and transform information in memory, to be able to translate mental conceptions into actions and be motivated to do so. Schunk and Zimmerman (1997) elaborated Bandura's theory into a social cognitive model of self-regulation in reading and writing. According to this model, the learning of a new task occurs in four phases: observation, emulation, self-control and self-regulation. Observation is a first step in the learning process. Then the learner emulates the model's general style. Encouraged by feedback, the learner can internalize the skill and finally use the strategy independently and in various contexts. Here we focus on the first phase: observation.

Observational learning is closely related to cognitive apprenticeship. In cognitive apprenticeship, an expert model verbalizes thought processes. Collins et al. (1989) describe how observation, as an element of cognitive apprenticeship, provides strategic knowledge to the learner and changes students' understanding of the modelled skill. For example, in writing students may not realize that experts organize their ideas, elaborate their goals and think about their audience. Observation may enhance this awareness about the task. Couzijn (1999), Braaksma et al. (2002) and Raedts et al. (2007) found that observation is effective for writing argumentative or synthesis texts. In these studies a multimedia learning environment was used; students watched videos of peer models performing a writing task while thinking aloud. Students who observed peer models performing a learning-to-write task wrote better texts afterwards than students who practised this learning-to-write task. Rummel and Spada (2005) proved the effectiveness of observation for learning to collaborate in computer mediated settings, Schworm and Renkl (2007) found beneficial effects in the domain of argumentation and Van Steendam et al. (2010) for cooperative revision tasks. It seems that modelling is beneficial for various types of tasks.

Several elements may influence the effectiveness of observational learning, such as the competence level of the models. Zimmerman and Kitsantas (2002) found that college students who observed a coping model who gradually improved her writing technique on a sentence-combining task did better than students who had observed a mastery model. In Braaksma et al.'s study (2002), the models in the videos performed short tasks about argumentation structures in writing. All students watched the same videos, but Braaksma asked students either to focus on the weaker model or on the more competent model of the pair ('which model did best and why?' vs. 'who did worst, and why?'). Evidence in support of the similarity hypothesis was found: when confronted with a new task, weaker writers learned more from focusing on the weaker model of a pair, while better writers learned more from focusing on the more competent model.

Observation should include evaluation. Braaksma, Rijlaarsdam, Van den Bergh and Van Hout-Wolters (2006) analysed students' observation processes and found that evaluation and elaboration are essential for the effectiveness of learning from observation. Sonnenschein and Whitehurst (1984) studied the effect of observation and evaluation compared to observation only for preschool children who acquire communication skills. The observation-evaluation condition performed better on speaking and listening tasks than the observation only condition. The additional evaluation task explains the transfer effects on listening and speaking according to Sonnenschein and Whitehurst. They describe evaluation skills as 'superordinate' skills since these skills transferred to speaking and listening, whereas increased performance on speaking and listening tasks did not transfer to evaluation skills. It seems advisable then to stimulate students to evaluate models and to elaborate on the models' behaviour after observation.

All in all it appears that observational learning is an effective approach in various domains. Attributes of the model (such as initial performance level) and students' thinking activities (evaluation and elaboration during and after observation) may influence the effect of observational learning.

1.2 Creative tasks and modelling

From the preceding section it can be concluded that observational learning is effective for structured domains such as mathematics as well as for ill-structured domains such as writing. The question arises whether observation can also be effective for creative tasks, which involve divergent thinking skills. High performance on artistic creative tasks requires original and novel responses. This means that the problem space in creative tasks is large; there are many possible solutions. Artists even have to discover their own task, the artistic problem, before they can start solving it (for example, finding out what to draw) (Getzels & Csikszentmihalyi, 1976). Therefore, creative tasks are extremely ill-defined.

Few studies have focused on modelling in a domain which requires students to formulate their own original problem. One explanation may be that using observation of models to enhance students' creativity may seem paradoxical. Creative work involves the generation of original ideas, while observing models may lead to imitation of products (the 'conformity effect'; Finke, Ward & Smith, 1992). However, observation of cognitive models is directed at developing a clear idea of how a task can be performed as demonstrated by Braaksma, Rijlaarsdam, Van den Bergh, and Van Hout-Wolters (2004). In their study the observation of peer models affected students' writing processes; the students who had learned to write by observing engaged in metacognitive activities during writing, such as planning, analysing and goal-orientation more often than students who had learned by practising writing. Therefore, we expect that the observation of someone who is thinking aloud while engaged in creative work affects the observer's future activities.

Studies in the area of worked examples have examined the effect of examples for learning in ill-defined domains (e.g. Rourke & Sweller, 2009; Van Gog, Paas, & Van Merriënboer, 2004; 2006; 2008). Worked examples differ from modelling examples (e.g. observational learning), since they involve 'ideal' problem solution steps presented as text. The underlying mechanisms, however, seem to be similar: students learn new procedures for problem solving and abstract general rules from the examples (Van Gog & Rummel, 2010). Rourke and Sweller (2009) found that students who studied worked examples of a task about recognizing designers' styles perform better than students who practised this task themselves. They concluded that process examples are as effective in ill-defined domains as they are in well-structured domains.

But what kind of knowledge should students acquire from observing and evaluating models in ill-defined domains? Hilbert, Renkl, Kessler, and Reiss (2008) introduced heuristic (worked) examples for ill-defined tasks, which demonstrate heuristic steps towards a solution. They studied the effect of these examples on mathematical proving skills, which include discovery behaviour. Heuristic knowledge was presented explicitly and self explanation prompts were directed at the heuristic level. This approach proved to foster learning. Van Gog et al. (2004, 2006, 2008) argue that experts' 'how' and 'why' process information enables students to deepen their understanding of solution procedures in ill-structured domains. For tasks with large problem spaces, learners need strategies to narrow the search space and select the most promising solution procedures. Therefore, students need to know why certain solution steps are taken. Van Gog et al. (2008) show that process information is indeed effective in the first phase of learning in electrical circuit troubleshooting.

Few studies examined the effect of modelling examples and artistic creative tasks. Teyken (1988) examined the effects of focused reflection on creative design. He incorporated observational learning tasks in an experimental curriculum for student art teachers. Students watched videos of designers at work. Teyken found that the students' design processes changed as a result of focused reflection, although the quality of their design products was not higher in the experimental condition than in the control condition. Observation was examined as part of an experimental curriculum, therefore it remains unclear which learning activity in the curriculum caused the effects measured. Anderson and Yates (1999) examined the effect of modelling on young children's clay works. They found that the quality of the clay works produced after modelling was higher than the quality of the clay works produced under regular conditions.

To conclude: few experimental studies have examined the effect of observation on artistic creativity and no studies have examined the effect of peer modelling in arts education for secondary school students. Therefore, in the present study we focus on creative process modelling through observational learning with peer models.

1.3 Creative processes

Which creative processes should be modelled in observational learning videos? Amabile (1996) proposed a componential model that encompasses three basic components necessary for creativity: (1) domain-relevant skills, (2) creativity-relevant skills and (3) intrinsic task motivation. Domain-relevant skills include basic skills (factual knowledge, technical skills, talent) relevant in a given domain. Creativity-relevant skills refer to an appropriate cognitive style and knowledge of heuristics for generating novel ideas. Intrinsic task motivation refers to motivation related to the task. Based on this model, we assume that observational learning videos should include heuristic strategies and cognitive style independent from the artistic domain. Subsequently, these heuristics are applied to the artistic domains: verbal (poetry writing) and visual (collage making).

Based on the literature available, we choose four activities that are relevant in creative processes to be modelled in observational learning videos: (1) initial and ongoing problem finding activities, (2) generating (large, deep) strings of ideas, (3) exploring generated ideas and (4) critically evaluating product (see method section and Table 1 for operationalisations).

Traditionally the creative process is described as consisting of four stages; preparation, incubation, illumination and verification (Wallas, 1926). During preparation the creator absorbs information and engages in problem finding and definition. During incubation the person is taking a step away from the creative process. During illumination a solution or great idea suddenly comes to mind and during verification the final product is created and edited. However, from recent research on artists and designers at work, it appears that creative processes co-occur throughout the work recursively. Finke, Ward, and Smith (1992) proposed a model of creative cognition called 'geneplore', emphasizing generative and exploratory cognitive processes. Generative processes involve the initial creation of an idea (or in fact a premature idea or 'pre-inventive structure'), whereas in the exploratory processes this pre-inventive structure is examined and interpreted in various ways (examined for emergent properties and implications). After the exploratory stage, pre-inventive structures may be refined and the process may repeat until a final product/idea has been developed. These two processes interact in cyclical sequences leading to creative products.

Runco (2003) describes creativity as the interaction of divergent (generative) and convergent (evaluative, critical) processes. Generating many ideas and many different types of ideas is called ‘divergent thinking’. Divergent thinking has been linked to creativity: those who produce many, diverse and original ideas are thought to be more creative. Divergent thinking can be stimulated through various brainstorming techniques, based on the idea that the presence of more ideas implies ‘more original ideas’ since original ideas are remote, at the end of a chain of associations (Mednick, 1962). Creativity also involves convergent processes, making choices from the wealth of options generated.

Getzels and Csikszentmihalyi (1976) observed that initial problem finding and on-going exploration throughout the work process appear to be important characteristics of creative production. They observed fine art students’ still-life drawing activities under experimental conditions (think alouds and videotapes). First the students had to compose a still life arrangement before drawing it. The students who engaged in an extended problem finding process, exploring many of the still-life objects in detail, produced work that was evaluated as more creative and original than that of the students who quickly took some objects and started drawing. Problem finding takes place in the preparation stage, but also during the production or editing phases (Getzels & Csikszentmihalyi, 1976). The students who produced more original work kept on exploring (for example by sketching) and re-defining their artistic problem (operationalized as many changes to the work-in-progress), whereas the students who produced less creative work hardly changed their initial idea of the final product.

1.4 Research questions

The aim of the present study is to develop and test a learning arrangement for creative tasks based on the principles of observational learning. We include two artistic domains in the study for reasons of generalizability (visual art, verbal arts). The control condition consisted of learning by practising.

The research questions are:

- Is observational learning more effective than learning by practising for students’ creative products, creative processes and motivation (intrinsic motivation, task value and self-efficacy)?
- Does ‘model similarity’ influence the effect of observational learning?

To test the model similarity hypothesis, two observation conditions were included; observation with a focus on a relatively competent model (observation^{strong}model condition) and observation with a focus on a less-competent model (observation-^{weak}model condition). Participants in both conditions watched the same videos with pairs of models engaged in creative work: one competent model and one weak model. What we varied was the focus during evaluation and elaboration; students were asked to focus either on the weaker or on the more competent model of the pair.

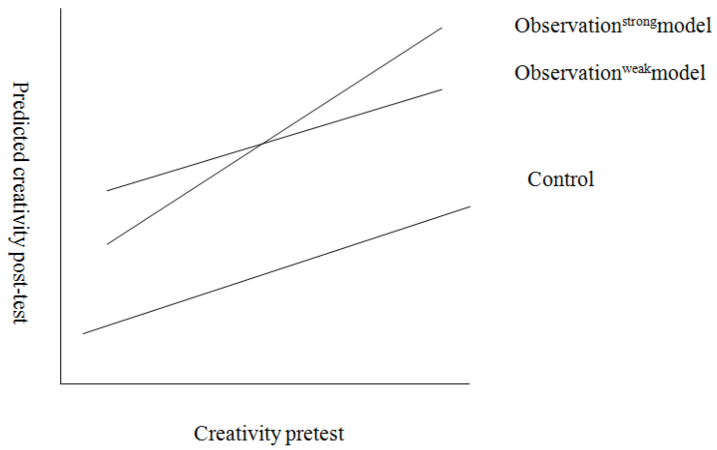
1.5 Hypotheses

We formulated the following hypotheses:

1) We expect observational learning to be more effective than practising. Concerning student products, we expect that: participants in the observational learning conditions will produce more creative collages and poems than students in the control condition.
2) We assume that observation will lead to more problem finding, generative, exploratory and evaluative processes, indicated by longer processes and more reprocessing. So we expect that: participants in the observational learning conditions will spend more time and revise more than students in the control condition.
3) We expect effects on motivation (intrinsic motivation, task value, self-efficacy), since Amabile (1996) considers motivation an essential component of creative skills. We expect that: Participants in the observation conditions will show higher levels of intrinsic motivation, task value and self-efficacy than students in the control condition.
4) Considering the model similarity hypothesis, we expect the similarity between the competence level of the model and the observer to influence the effect of observational learning; students with a low pre-test score will profit more from the condition with a focus on the weaker model, while students with a high pre-test score will benefit more from focusing on competent models. This means that the treatment effect differs for students with different ability. We expect that: treatment effects will interact differently with learner characteristics in the two observation conditions: the higher the scores on the pre-test, the more beneficial the observation^{strong}model condition (OSM); the lower the pre-test scores, the more beneficial the observation^{weak}model condition (OWM).

Figure 1 illustrates the results expected on product creativity for the different conditions based on students' pre-test scores.

Figure 1. Expected main effect and interaction effect.



2. METHOD

2.1 Research design

We conducted two experiments successively, one for poetry writing and one for collage making. A pre- and post-test control group design with three conditions was implemented: learning by observation with a focus on a relatively competent model (observation^{strong}model, OSM), learning by observation with a focus on a weak model (observation^{weak}model, OWM) and learning by doing (Practice). Students in the observation conditions observed the same videos, each showing two models in action, the difference being the instructional focus during evaluation and elaboration of model behaviour. Students from each participating class were randomly assigned to one of the three conditions; they stayed in the same condition for both experiments (poetry writing and collage making).

2.2 Participants

Three schools, with two classes each, from different regions in the Netherlands participated. Five teachers volunteered to take part in the experiments. Since the students were randomly assigned to the conditions, one teacher was responsible for all the three conditions within the class. The participants were 153 fifteen to sixteen – years old secondary school students (10th grade, pre university and higher general secondary education; 56 boys and 97 girls). 22 Students were excluded from the dataset, because they did not attend all the lessons. In the final data set, there were 44 students who had participated in the OWM condition, 42 in the OSM condition, and 45 in the control condition. The students participated during their regular 'Cultural and Artistic Education'[4] classes.

2.3 Observational learning materials

We made observational learning videos focusing on various phases of the creative process. In the collage videos, the collage-in-progress and the model's hands were visible on the screen, while learners heard the model's thinking aloud as a voice over. In the poetry videos, the 'poem-under-construction' was seen on the screen and the voice of the student who was thinking aloud could be heard. We selected four domain independent processes and applied them to both domains: problem finding, generating, exploring and evaluating ideas. In Table 1, we present the content of the observational learning videos for both domains. As shown in the table, the videos contained heuristic information, process information ('how' and 'why' information) and attitudinal information (flexibility and motivation). Scripts for the videos were based on actual students' processes, then written in scenario format and role played from the scenarios. We speeded up some parts of the actual student processes or exaggerated a little to make them more attractive learning materials.

During the intervention, the students in the observation conditions did not engage in collage making or poetry writing. Instead, we asked them to watch the videos on the computer and to make notes. To direct students' attention to the relevant processes and evoke a comparison process, contrasting approaches (rather weak and rather strong) were shown in the videos by two different models. As shown in Table 1, students watched both relatively competent and relatively weak peer models at work.

The models in the videos were not labelled as 'competent' or 'weak', but after the students had watched each video, we asked them to compare, evaluate and elaborate on the behaviour of the models by answering a question. Following Braaksma et al. (2002), we asked the students in the OSM condition: '*Which student did better in your opinion? Explain: what did this student do so well?*' The students in the OWM condition considered the question: '*Which student did less well in your opinion? Explain: what did this student do not so well?*' Most of the students were able to point out which model did better and which model did not do so well: on average 89% of the students (*SD* = 5.58) answered the questions as intended. This means that the students adhered to the instruction (and thus to the conditions they were assigned to). The students could replay the videos if they wanted.

In the control condition, students engaged in collage making and poetry writing, without watching models. The tasks were the same as those shown in the observational learning videos. To ensure that only modelling of the work *process* made a difference between the conditions and not product modelling, students in the control conditions reflected on the same final products as those that were shown in the ob-

[4] *CKV is a compulsory subject in Dutch upper secondary education. It includes: visual arts, music, dance, theatre, literature, architecture etcetera. The focus is on perception and reception, but CKV also includes a productive component.*

servational learning videos. We made sure that the total amount of time spent remained the same in both conditions. Students spent one lesson hour on poetry writing and one lesson hour on collage making.

Table 1. Content of the observational learning videos

Video	Domain independent Processes	Content of video: Collage	Content of video: Poetry
1	Problem finding vs. Routine solution.	Problem finding, generating various ideas before actual collage making vs. fixation on one first idea (4.23 min.).	Brainstorming and exploring before actual writing vs. directly writing a final version on the basis of first associations (4.09 min.).
2	Generating and exploring many ideas (content) vs. Generating just one idea.	Trying out several ideas, evaluating vs. immediately carrying out first idea without evaluation (5.19 min.).	Using items from the brainstorm for further exploration, forming poetry lines vs. direct copying of the brainstorm into final text (3.23 min.).
3	Exploration of form vs. Blocking.	Solving problems that arise with flexibility by exploration of chosen idea in several visualizations, exploring the material, vs. inflexible approach, inability to solve a problem in a satisfying way (5.21 min.).	Rereading and exploring written text (sound, meaning) vs. fixation on rhyme (3.42 min).
4	Critical evaluation and revision vs. No evaluation.	Evaluation and revision of essential visual elements vs. revision on very detailed, less relevant level (5.46 min.).	Rereading, evaluation and revision on large scale vs. no rereading, revision of spelling and punctuation (4.29 min.).

2.4 Measures

The pre- and post-tests consisted of collage and poetry assignments. Amabile (1982) showed that collage tasks are very suitable for measuring creative behaviour since few technical skills are required. At the pre-test, students were asked to make a collage of a human figure, consisting of geometrical shapes (30 minutes). At the post-test, the students were asked to make a collage of a 'strange creature' (60 minutes).

For both tests, all the students received an identical set of materials (scissors, glue, coloured paper, a set of magazines). The students were instructed to be as creative as possible.

The poetry pre-test and post-test each consisted of two short tasks (10 minutes each). Since most students had never written poetry before, we used tasks that were short and structured (see appendix A). Since the tasks were short, it was feasible to administer two tests and therefore raise validity. Pre-test task 1 and post-test task 1 had been used in previous research (Broekkamp, Janssen, & Van den Bergh, 2009; Ruscio, Whitney, & Amabile, 1998). The students were instructed to be as creative as possible.

Art students from the Amsterdam School of the Arts (collages) and university students in linguistics (poems) scored all the products holistically for creative performance on a 0-200 scale with the support of anchor products. These anchors were based on scores from other raters in previous scoring procedures (among others Broekkamp et al., 2009) and illustrated a non-creative, medium creative and highly creative product with fixed scores of 50,100, 150 (see Appendix B and C). First we discussed the degree of creativity of the anchor products with the raters. Then the raters spread the products over three piles (non-creative, medium creative and highly creative) using the anchor products. Subsequently, the raters went through each pile separately, assigning a score to each product. Jury reliability for the collage tests was sufficient (pre-test .77, post-test .72). For the poems we used a design of overlapping rating teams (Van den Bergh & Eiting, 1989), because of the large number of products and the time required to assess these. We obtained four scores per poem. The estimated reliability for the rating of the poems was sufficient (ρ= .91 and ρ = .74 for the pre-tests tasks and ρ = .83 and ρ = .74 for post-tests tasks).

It was hypothesized that observation would lead to more problem finding, generative, exploratory and evaluative processes. Since the classroom setting did not allow for large scale think aloud protocol collection, we collected secondary process measures. We focused on process time and revision behaviour as we assumed that exploration and problem finding activities would result in longer work processes and more production. To collect traces of the collage making processes, we counted the number of images that students cut out of magazines but which they did not use in their final products. We regarded this as an indication of problem finding, exploration and revision in the processes. Since nearly all the students used all the available time, process time was not used as a process variable for collage making.

For the poetry writing process, we used Inputlog, a keystroke logging programme, to register students' processes (Leijten & Van Waes, 2005). Due to technical problems with the installation of the software at one school the processes of only two thirds of the students were recorded. From these keystroke logging data, we calculated the process times and revision ratios. Process time was defined as the total time spent on the task; revision ratio is the number of words in the final text as a percentage of the total number of words produced.

At pre-test and post-test, the Motivated Strategies for Learning Questionnaire (Pintrich, Garcia, & McKeachie, 1991) was administered, adapted to poetry writing and collage making. The questionnaire consisted of 19 items for each of the two artistic domains; 38 questions in total. Likert scales for students' perceptions on their intrinsic motivation (7 items), task value (5 items) and self-efficacy (7 items) were used. The reliability of the scales varied between .86 and .93. Additionally, to acquire information about the students' capacity level, a verbal IQ test (DAT, 1984) was administered (Cronbach's alpha .72).

2.5 Procedure

During the observation sessions, the students worked individually on the computer. For collage making, students in the control group worked in an adjacent room. During the pre- and post-test, the students worked in one large room, creating a collage individually. For poetry writing, the participants from all three conditions were present in the same classroom, guided by worksheets; either watching and evaluating videos or writing poems themselves. The students received a short instruction by the researcher, the teacher's role was that of an organizer, since the materials were largely self-explanatory.

The students worked for six sessions of about 60 minutes. First the students filled in a questionnaire on motivation, task value and self-efficacy and completed the pre-tests on poetry writing (session1). Then they did the pre-test on collage making and the first part of the verbal IQ test. The intervention (about 60 minutes) and the post-tests on poetry writing (about 20 minutes) were completed in session 3. Collage making was the content of the 4th session (about 60 minutes). During the fifth session the post-test on collage making was administered, and in the sixth session, the students filled in the post-test questionnaire on motivation and the second part of the verbal IQ test. There were minor differences in procedure between schools due to different class schedules.

2.6 Analyses

We cannot test the learning effect from pre-test to post-test, because the level of difficulty of the two tests is unknown and, therefore, we cannot compare the outcomes. Therefore, we can only examine the effect of condition on the post-test.

Figure 1 shows that we expected observational learning to be more effective than practising. It also shows that we expect weaker students to perform better in the OWM condition while stronger students would perform better in the OSM condition. Since we have multiple observations per individual (more than one test) a mixed model analysis was chosen, in which both the variance *within* students and *between* students are estimated simultaneously. Several models were tested. In the first model, the so-called empty model, we estimated an intercept (next to both variances). In the second model we added differences between conditions, whereas in the third model we took initial differences into account as well. Hence, in this model we assume that the relation between pre-test and post-test would not vary between conditions. This assumption is relaxed in the fourth model, which allows differences in means between conditions as well as differences in regression from post-test on pre-test. The difference in fit between these (nested) models can be tested by comparing -2loglikelihood, as the difference in –2loglikelihood in nested models is χ2-distributed (with the difference in number of estimated parameters as degrees of freedom).

For poetry writing two learner variables (pre-test level of creativity in poetry writing and verbal IQ) were entered successively into the analysis. So, six models were tested; (0) intercept only, (1) a model without a covariate, (2) a model with pre-test level as a covariate, (3) a model with pre-test level as a covariate and allowing its effect to vary between the conditions, (4) a model with two covariates and (5) a model allowing the influence of the second covariate to vary between the conditions. Subsequently these models were applied to measure effects on products, processes, motivation, task value and self-efficacy.

3. RESULTS

We will first report on the effect of observation on the creativity of the students' products (collages and poems). Then, we will report on the process effects, followed by the results for intrinsic motivation, task value and self-efficacy.

3.1 Creativity of the collages

Table 2 provides the mean scores for the students' collages, at pre-test and post-test, for the three conditions. Pre-test scores did not differ significantly (*p* = .752) between the conditions.

Table 2. Mean scores collages, z-scores (rating scale 0-200)

	Pre-test		Post-test	
	M	*SD*	*M*	*SD*
Observation	-.05	1.01	.18	1.03
Observation^{strong}model	-.04	.98	.06	.99
Observation^{weak}model	-.06	1.04	.29	1.06
Control group	.09	.99	-.31	.88

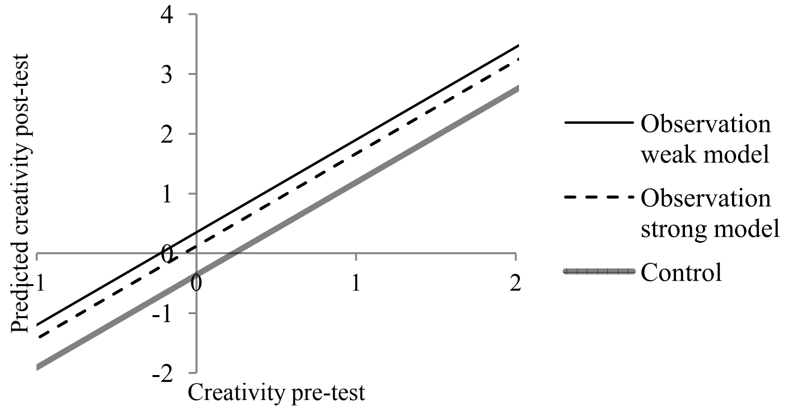
In Table 3 the five models analysing the effect of observation on the creativity of the collages are presented. Model 1 (distinguishing between conditions) fits the data significantly better than model 0 (Intercept only). This is shown in the right hand column of Table 3 (*p*=.02). Therefore, we reject model 0 in favour of model 1. Subsequently, model 2, including pre-test as a covariate, fits better than model 1 (*p*=<.001), so we must reject model 1 in favour of model 2. Since the difference between -2LL of model 3 and -2LL of model 2 is not significant (*p*=.11), we must choose model 2. According to this model, the effect of the pre-test does not differ between the conditions. This implies that, in contrast to our expectation as presented in figure 1, the regression lines for the predicted post-test scores are parallel (Figure 2). In other words; the slopes of the regression lines do not differ significantly between the conditions: the influence of pre-test level is equal for the conditions. This means that we did not observe a model similarity effect.

Table 3. Comparison of models with creativity of collages as a dependent variable (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	331.0				
1	Condition as a factor	323.3	0 vs 2	7.7	2	.02
2	Condition as a factor and pre-test as a covariate	303.7	1 vs 2	19.6	1	< .001
3	Condition as a factor and the effect of pre-test differs between the conditions	301.2	2 vs 3	2.5	2	.11

Mixed models analyses with pre-test as a covariate revealed that both observation conditions performed significantly better than the control condition (OSM vs. practice: mean difference= .480, *se*=.226, *p*=.036, OWM vs. practice: mean difference= .707, *se*=.221, *p*=.002). This means that at post-test the students who had observed made significantly more creative collages than the students in the control condition. The students in the OSM condition performed .48 standard deviations better than the students in the control condition and students in the OWM condition performed .71 standard deviation better than students in the control condition. The OSM condition and OWM condition did not differ significantly from each other (*p*=.328).

Figure 2. Regression lines for three conditions with z-score pre-test predicting post-test score on collage creativity.



3.2 Creativity of the poems

In Table 4 the indices for the poems at pre-test and post-test per condition are presented. At the pre-test, the creativity of the students' poems did not differ significantly between conditions (p =.369).

Table 4. Mean scores of students' poems, z-scores

	Pre-test1		Pre-test2		Post-test1		Post-test2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Observation	.12	.94	.08	.82	.09	.78	.03	.82
Observation^{strong}model	.10	.83	.04	.68	.03	.78	-.20	.68
Observation^{weak}model	.15	1.06	.12	.96	.14	.79	.24	.89
Control group	-.19	.71	-.12	.69	-.10	.70	-.09	.71

In Table 5 we present and compare five models for the effect of observation on the creativity of the poems. As shown in this table, model 1 (distinguishing between conditions) fits the data better than model 0 (only intercept). Therefore, we continued to compare model 1 and 2 (including pre-test as a covariate). As -2LL is significantly smaller for model 2, we concluded that pre-test is a significant covariate and that we therefore must reject model 1 in favour of model 2. Subsequently we allowed the influence of the pre-test to vary between the conditions (model 3). This did not lead to a significant improvement of the model, so we rejected model 3. Model 2 was compared to model 4 which takes the second covariate (verbal IQ) into account. Model 4 fits the data better than model 2. Finally we compared model 4 to a fifth model, allowing the influence of verbal IQ to vary between the conditions, which did not lead to a significant improvement of the model. Therefore, model 4 was chosen.

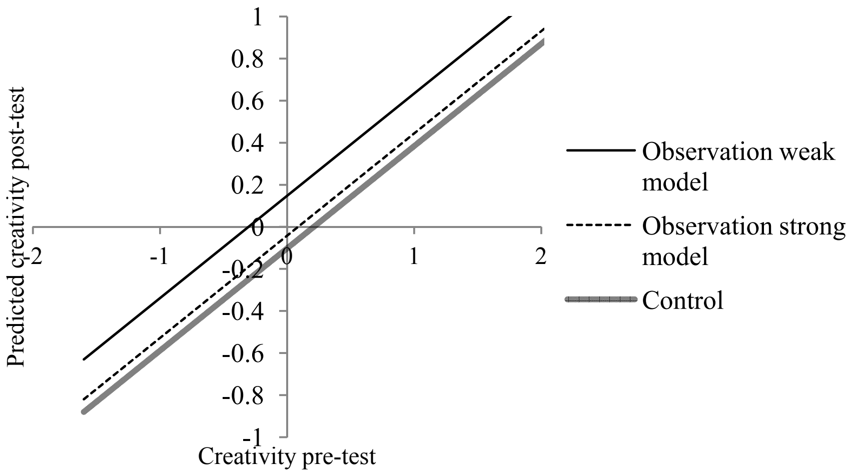
Table 5. Comparison of models with creativity in poetry writing as a dependent variable (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	588.9				
1	Condition as a factor	583.0	0 vs. 1	5.9	2	.05
2	Condition as a factor and pre-test as a covariate	501.7	1 vs. 2	81.3	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions.	498.8	2 vs. 3	2.9	2	.23
4	Condition as a factor and pre-test and IQ as covariates	478.4	2 vs. 4	23.3	1	<.001
5	Condition as a factor and pre-test as a covariate and the effect of verbal IQ differs between the conditions	476.4	4 vs. 5	2.0	2	.37

Pre-test and verbal IQ proved to be valid covariates, but in contrast to our expectation, their effect did not differ between the conditions. This means that the regression lines of the predicted post-test scores on the pre-test scores are parallel (Figure 3), which implies that there is no model similarity effect.

According to our analysis, the students in the OWM condition performed significantly better than the students in the OSM condition (mean difference=.25, *se*= .12, *p*= .034), but not better than the students in the control condition (*p*=.109). The control condition and OSM condition did not differ significantly (*p*=.591). This means that the students in the OWM condition wrote the most creative poems.

Figure 3. Regression lines for three conditions with z-score pre-test predicting post-test score on poem creativity.



3.3 Collage making processes

For the collage task we examined whether the students who had observed revised more than the students in the control group. Table 6 presents the mean number of images that were cut out, but were not used in the final product. No difference between conditions on the pre-test (*p* =.475) was observed.

Table 6. Collage processes, means of unused images, z-scores

	Pre-test		Post-test	
	M	*SD*	*M*	*SD*
Observation	.06	1.21	.18	1.15
Observation^{strong}model condition	.17	1.55	.13	1.01
Observation^{weak}model condition	-.05	.75	.23	1.28
Control	-.10	.43	-.30	.57

The same models as used for the collage product analyses were tested (See appendix D1 for a table comparing the different models). The model with pre-test as a covariate was found to fit the data best (model 2). The influence of the pre-test did not vary between the conditions, resulting in regression lines running parallel.

The students in the OWM condition had more unused shapes than the students in the control condition (mean difference= .532, *se*=.222, *p*=.018), while students in the OSM condition did not have significantly more unused images than students in the control condition (*p*=.192). The OSM and OWM condition did not differ significantly in this respect (*p*=.318). This indicates that the students in OWM revised significantly (.53 standard deviation) more than the students in the control condition.

3.4 Poetry writing processes

For poetry writing, we examined two process variables (process time and revision ratio) for differences between the conditions (see Table 7). At pre-test, the conditions did not differ significantly in process time (*p*=.410, *p* = .387) or revision ratio (*p* = 385, *p* =.638).

Table 7. Poetry writing processes, means of process time and revision ratio (words in final text/words produced), z-scores

		Pre-test1		Pre-test2		Post-test1		Post-test2	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Process time	Obs^{strong}model	-.19	.96	.05	1.02	-.11	1.07	-.01	.86
	Obs^{weak}model	.07	1.08	.15	.98	.59	.84	.12	1.09
	Control	.13	.97	-.19	1.00	-.49	.78	-.11	1.04
Revision ratio	Obs^{strong}model	.19	.98	-.06	1.11	-.13	.77	-.22	.93
	Obs^{weak}model	-.18	1.13	-.09	.92	-.32	1.38	-.14	1.00
	Control	-.05	.91	.14	.99	.44	.46	.33	1.01

For both process variables the fourth model was found to fit the data best: the model with pre-test and verbal IQ as covariates (see appendix D2 and D3). The effect of the covariates did not differ between the conditions. So again, parallel regression lines must be assumed. For process time it was found that OWM differed significantly from the control condition (mean difference=.529, *se*=.196, *p*=.008), but OSM did not differ significantly from the control condition (*p*=.475). Also OWM and OSM differed significantly (mean difference= .392, *se*= .195, *p*=.047). This means that students in the OWM condition had the longest process times of the students in all the three conditions.

For revision OWM and OSM differed significantly from the control condition (mean difference= -.576, *se*=.186, *p*=.003 and mean difference = -.567, *se*= .181, *p*=.002). The students from both observation conditions revised significantly more at post-test than the students in the control group. The observation conditions did not differ significantly from each other in this respect (*p*=.961): the students who observed revised more than the students who did not observe. We conclude that the writing processes of the students who observed differed from the processes of the students who did not observe.

3.5 Motivation, task value and self-efficacy

Table 8 presents the mean scores on intrinsic motivation, task value and self-efficacy. At pre-test, there were no significant differences between the conditions (collage making: intrinsic motivation: p = .736; task value: p = .302; self-efficacy: p = .997; poetry writing: intrinsic motivation: p = .459; task value: p = .484; self-efficacy: p = .715).

For collage making, the model with pre-test as a covariate fits best (see appendix D4-D6) for intrinsic motivation, task value and self-efficacy. We found that for intrinsic motivation and self-efficacy, the students in OWM had significantly higher scores than the students from the control condition (intrinsic motivation: mean difference = .528, se = .209, p = .013; self-efficacy: mean difference = .573, *se* = .205, p = .006). The students in OSM did not differ significantly from the students in the control condition (intrinsic motivation: p = .134; self-efficacy: p = .129). Scores between both observation conditions did not differ significantly either (intrinsic motivation: p = .338; self-efficacy: p = .234). For task value, both observation conditions differed significantly from the control condition (OSM: mean difference = .446, *se* = .208, p = .034; OWM: mean difference = .528, *se* = 203, p = .010), but not from each other (p = .693). Task value proved to be higher for the students who observed than for the students who did not. Only the students who focused on the weaker model had higher intrinsic motivation and self-efficacy scores.

Table 8. Observed pre-test and post-test means, z-scores: intrinsic motivation, task value and self-efficacy for collage making and poetry writing (1: strongly disagree, 7: strongly agree)

Variable	Condition	Collages				Poems			
		Pre-test		Post-test		Pre-test		Post-test	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Intrinsic motivation	$Obs^{strong}model$	.09	1.06	.05	1.02	.13	.97	.15	.90
	$Obs^{weak}model$	.01	.96	.19	.79	.02	.93	.09	.88
	Control	-.09	.99	-.24	1.13	-.14	1.09	-.25	1.17
Task Value	$Obs^{strong}model$	-.07	1.04	.11	1.05	.13	1.00	.24	.95
	$Obs^{weak}model$	.22	.98	.15	.87	-.14	.88	-.10	.91
	Control	-.12	.97	-.28	1.02	.01	1.10	-.17	1.09
Self - efficacy	$Obs^{strong}model$	.01	.99	.03	1.06	.11	1.14	.18	1.02
	$Obs^{weak}model$	-.01	1.00	.22	.81	-.05	.84	-.05	.85
	Control	-.00	1.03	-.25	1.08	-.05	1.01	-.15	1.09

For poetry writing, the model with pre-test and verbal IQ as covariates (see appendix D7-D9) was applicable for intrinsic motivation and task value. For self-efficacy, condition as a factor did not improve the model significantly. In other words, we observed no effects of condition on self-efficacy. Closer examination of the results on intrinsic motivation and task value did not reveal any significant difference between the conditions (intrinsic motivation: OSM vs. control: p = .250, OWM vs. control: p = .562; task value: OSM vs. control: p = .241; OWM vs. control: p = .287). For poetry writing, the students who observed were not more motivated, nor did they have higher task value or self-efficacy scores than the students who did not observe.

4. DISCUSSION

We investigated the effect of observational learning on students' performance on creative tasks in the visual and verbal domain. We aimed at answering two questions: is observational learning more beneficial than learning by practising for creative products, processes and motivation (intrinsic motivation, task value and self-efficacy)? And does the 'model similarity effect' influence the effect of observational learning? We expected the students who observed to produce more creative products than the students who did not observe. Concerning the processes, we hypothesized that observation would result in longer session times and more revision than learning by practice We expected intrinsic motivation, task value and self-efficacy to be higher after observation. Regarding model similarity, we hypothesized that the effect of pre-test and verbal IQ (in the case of poetry writing) would differ between the conditions.

For collage making, our hypotheses for main effects were largely confirmed. The students who observed others at work produced collages that were rated as more creative than the students who learned by practising. The students' processes of collage making differed between conditions at the post-test: the students revised more (in terms of more unused but cut out shapes) in the observation^{weak}model (OWM) condition than in the control condition. This difference was not significant for the observation^{strong}model (OSM) condition. For OWM, intrinsic motivation, task value and self-efficacy were significantly higher after the intervention than for the students in the control group. For OSM, this was only the case for task value. For none of the variables did we find evidence in support of the model similarity hypothesis.

For poetry writing we found that OWM resulted in better poems than OSM, although the observation conditions did not score significantly better than the control condition. For processes, OWM resulted in longer processes than the control condition, while both observation conditions resulted in more revision than the control condition. No effects were observed on intrinsic motivation, task value and self-efficacy. We conclude that the hypotheses regarding effects on products and motivational variables must be rejected, but that poetry writing processes were affected significantly by observation. For none of the variables did we find evidence in favour of the model similarity hypothesis.

In general, we found indications that the students who focused on the weaker model performed better in both domains (products and processes). Initially we did not expect differences in the mean performance between the two observation conditions. Instead, we expected different students to behave differently in the conditions. Possibly, the task of evaluating the videos of the relatively more competent model was too difficult for the students. The stronger model might have served as a frame of reference when the students had to explain why the weaker model did worse (Braaksma et al., 2002), which is easier than explaining what a good model does well. Creative processes of the majority of the students might have been more similar to those of the weaker student in the poetry videos, since Dutch students generally have little experience in poetry writing. Possibly a 3 x 2 design with condition (practice; observational learning from weak model only; observational learning from strong model only) and student ability (weak versus strong) as factors would have provided more information on differential effects of student ability.

Some validity concerns about the present study must be discussed. One limitation concerns the measurement of creativity of products. We did not score the products on technical qualities and therefore we can only assume that the judges were able to distinguish between creative performance and technical quality. Although raters seemed to be able to distinguish levels of creativity, validity might be warranted in future studies by adding non-creative dimensions to the rating procedure as well.

Another concern is the measurement of processes. The process measures we used are only indirectly related to cognitive processes. For example, the ratio of words in the final text and the number of words produced does not distinguish between prewriting activities, ongoing exploration and actual revision behaviour. However, our measures proved to be sensitive enough for detecting condition effects. Currently, the setting in real classrooms prevented us from measuring processes in a more direct way (e.g. by using think aloud methodology). In future studies process measurement could be optimized, attempting to get closer access to the actual processes. An interesting procedure seems to be a secondary task procedure as installed in classroom settings in writing process studies (Torrance, Fidalgo, & Garcia, 2007) which can be adapted to visual tasks.

Another issue is that the procedures in the two experiments (collage making and poetry writing) were different in time and order. The poetry experiment was the first to take place, so this may have affected the collage experiment and its outcomes. Moreover, for practical reasons the duration of the tests and the test moments differed between experiments, which hampers generalizability across domains. More convincing results were found for the visual domain. Possibly using a visual medium (video) is more effective in a visual domain, collage making, than in a non-visual domain such as poetry writing. More research is needed to clarify this issue.

Furthermore, external validity is an issue which deserves attention. External validity can be promoted by implementing several versions of a treatment. In the current experiment we implemented two versions of a treatment (focus on weak and strong model). In future studies diverse treatment operationalisations could be further explored. Using different operationalisations may complicate the interpretation of the results, as we have seen, but generalizability would benefit from it. To satisfy experimental requirements, the setting in the current study was not completely realistic. The students had little freedom; they had to finish their creative products in one session, within a time limit. In a more realistic context, they might have opportunities to leave the products for some time and return to them for further elaboration at later moments. The control condition consisted of practice guided by a workbook, without teacher involvement. In art education, however, the teacher usually plays a crucial role. In future studies, it will be interesting to study the effect of observation in real educational practice. Implementing observational learning in actual art classes requires embedding observational learning videos in a lesson series (accompanied by feedback and practice).

All in all, the results suggest that it is worthwhile to implement observational learning in arts education. At least for visual arts we have demonstrated positive effects. We certainly do not suggest replacing artistic practice by observation as we did in the experiment. Instead, we recommend enriching the arts classes with observational learning. From a theoretical perspective, it can be concluded that observational learning is an effective learning tool, not only for structured tasks, but also for ill-defined creative tasks in arts education. Even when students are asked to produce original work, modelling examples may support them. In addition, we think that this study shows a new direction of studying interventions that foster creativity, taking both effects on process and product into account.

ACKNOWLEDGEMENTS

This study was carried out with a grant from The Research Institute of Child Development and Education of the University of Amsterdam. We would like to thank the anonymous reviewers for their thorough reading of the manuscript and their feedback.

APPENDIX A: POETRY TASKS

Pre-test 1:
Write a cinquain that starts with the word 'snow'.
This is the form of a cinquain:

Line 1: first word (given)
Line 2: two adjectives about the first word.
Line 3: three verbs about the first word
Line 4: a sentence about the first word (you may decide about the length)
Line 5: repetition of the first word

Available time: 10 minutes (an example was provided)

Pre-test 2:
Write a poem based on the following steps:

- Mention a place in the house
- Make up a line with a colour
- Make up a line with a domestic device and a sound
- Tell something about the weather
- Make up a line that is related to the lines above
- Repeat the first line
- Repeat the second line with a small change

Available time: 10 minutes (an example was provided)

Post-test 1:
Write a poem of five lines that contains the following words:
Music, bike, shiver, green, resembles

Each line should contain one of these words (each word should only be used once)

Available time: 10 minutes.

Post-test 2:
Write an animal haiku about a lion.
A haiku contains 17 syllables, divided over three lines of the poem. This is the form of a haiku:

- Line 1: five syllables.
- Line 2: seven syllables.
- Line 3: five syllables.

Available time: 10 minutes

APPENDIX B: ANCHORS COLLAGE POST-TEST

Score 50

Score 100

Score 150

APPENDIX C

Anchors poetry post-test (These anchors are translated from Dutch)

Score 50

I went by bike to tennis class
The music is full of chatter
I feel a shiver over my back
The green is from the grass
Because it looks like a pool [in Dutch this rhymes with the previous line]

Score 100

I cycle through the forest
My ipod plays music
Leaves from the trees are luminously green
Suddenly there is a shiver
It seems to be winter

Score 150

Everything that seems normal to you,
Makes me shiver
The colour green
Hearing the music
That day we were together, you behind me on my bike

APPENDIX D

Comparison of models with process variables as dependent variables.

Table D1. Four models analysing effects on revision in collage making (-2LL)

	Model	-2LL	Models Compared	X^2	*df*	*p*
0	Intercept only	331.0				
1	Condition as a factor	324.3	0 vs. 1	6.7	2	<. 03
2	Condition as a factor and pre-test as a covariate	305.1	1 vs. 2	19.2	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions	302.8	2 vs. 3	2.3	2	>.32

Table D2. Six models analysing effects on process time in poetry writing (-2LL)

	Model	-2LL	Models Compared	X^2	*df*	*p*
0	Intercept only	505.1				
1	Condition as a factor	494.0	0 vs. 1	11.1	2	<.004
2	Condition as a factor and pre-test as a covariate	470.5	1 vs. 2	23.5	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions.	469.7	2 vs. 3	.8	2	0.67
4	Condition as a factor and pre-test and IQ as covariates	457.8	2 vs. 4	12.7	1	<.001
5	Condition as a factor and pre-test as a covariate and the effect of verbal IQ differs between the conditions	457.8	4 vs. 5	0.0	2	1.0

Table D3. Six models analysing effects on revision in poetry writing (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	505.8				
1	Condition as a factor	492.1	0 vs. 1	13.7	2	0.001
2	Condition as a factor and pre-test as a co-variate	475.7	1 vs. 2	23.6	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions.	474.1	1 vs. 3	1.6	2	0.45
4	Condition as a factor and pre-test and IQ as covariates	464.9	2 vs. 4	9.2	1	0.002
5	Condition as a factor and pre-test as a co-variate and the effect of verbal IQ differs between the conditions	463.9	4 vs. 5	1.0	2	0.61

Table D4. Four models to analyse effects on intrinsic motivation in collage making (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	365.1				
1	Condition as a factor	358.6	0 vs. 1	6.5	2	0.038
2	Condition as a factor and pre-test as a covariate	334.8	1 vs. 2	23.8	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions	332.8	2 vs. 3	2.0	2	0.37

Table D5. Four models analysing effects on task value in collage making (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	365.1				
1	Condition as a factor	356.9	0 vs. 1	8.2	2	0.02
2	Condition as a factor and pre-test as a covariate	328.1	1 vs. 2	28.8	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions	327.3	2 vs. 3	.8	2	0.67

Table D6. Four models analysing effects on self-efficacy in collage making (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	365.1				
1	Condition as a factor	358.1	0 vs. 1	7.0	2	0.03
2	Condition as a factor and pre-test as a covariate	330.8	1 vs. 2	27.3	1	<.01
3	Condition as a factor and the effect of pre-test differs between the conditions	330.5	2 vs. 3	.3	2	0.86

Table D7. Six models analysing effects on intrinsic motivation in poetry writing(-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	365.1				
1	Condition as a factor	358.7	0 vs. 1	6.4	2	0.04
2	Condition as a factor and pre-test as a co-variate	331.0	1 vs. 2	27.7	1	<.001
3	Condition as a factor and the effect of pre-test differs between the conditions.	329.1	2 vs. 3	1.9	2	0.39
4	Condition as a factor and pre-test and IQ as covariates	315.6	2 vs. 4	13.5	1	<.001
5	Condition as a factor and pre-test as a co-variate and the effect of verbal IQ differs between the conditions	314.5	4 vs. 5	1.1	2	0.58

Table D8. Six models analysing effects on task value in poetry writing (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	365.1				
1	Condition as a factor	357.2	0 vs. 1	7.9	2	0.048
2	Condition as a factor and pre-test as a covariate	328.6	1 vs. 2	28.6	1	<.0.001
3	Condition as a factor and the effect of pre-test differs between the conditions.	327.8	2 vs. 3	.8	2	0.67
4	Condition as a factor and pre-test and IQ as covariates	314.1	2 vs. 4	13.7	1	<0.001
5	Condition as a factor and pre-test as a covariate and the effect of verbal IQ differs between the conditions	312.9	4 vs. 5	1.2	2	0.55

Table D9. Two models analysing effects on self-efficacy in poetry writing (-2LL)

	Model	-2LL	Models compared	X^2	*df*	*p*
0	Intercept only	362.2				
1	Condition as a factor	365.4	0 vs. 1	5.8	2	0.055