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

Groenendijk, T.

Publication date
2012

Chapter 5

DISCUSSION COMPARISON OF EXPERIMENTS AND SUGGESTIONS FOR FUTURE RESEARCH

1. INTRODUCTION

In this chapter we will summarize the main findings from the empirical studies. The two experiments, which are the central part of this thesis, will be compared with regard to the methodological and other decisions we made and with regard to validity issues. Next, directions for future research will be outlined.

1.1 Three studies: Main findings

In the study on poetry writing processes by secondary school students (chapter 2), we have shown that students' writing processes consist mainly of new text production. Contextual revision (revision within already written text) is infrequent. The processes of the students differ with regard to the distribution of writing activities over the process, amount of revision and linearity (the order of the lines in the final poem represents the order in which they are written, or not). 'Ideal' processes or processes that lead to good poems are preceded by much text production and many large scale contextual revisions towards the end of the process. This study on poetry writing processes provided us with useful information for designing interventions aimed at supporting artistic creativity.

We carried out two experimental studies to test the effect of observational learning as reported in chapter 3 and 4. In both experiments, we demonstrated the positive effects of observation for visual art production. In experiment 1, two experimental groups did better than the control group on creativity. In experiment 2, the experimental group did better than the comparison group by producing more creative designs. With regard to the creative processes we also demonstrated beneficial effects of observation in both experiments. In experiment 1, no model similarity effects, effects related to the similarity in competence between the model and the observer, were found. For some variables in experiment 2 differences in prior knowl-

edge (due to more or less experience with the visual arts subject in school) evoked different responses to the conditions.

In experiment 1 we also tested the effect of observation in the verbal domain (poetry writing). No effect of observation on the creativity of the poems was found, but students in both observation conditions showed different processes than students in the control condition: the students in the observation conditions revised significantly more than the students in the control condition. Possibly, it takes practice for recently learned strategies to result in good products.

1.2 Comparing two experiments

Central in the thesis are two experiments (chapter 3 and 4) testing the effect of observational learning. In Table 1 we present the main differences and similarities between these experiments. The experiments differed with regard to participants, setting, experimental design, treatment and measures. In the following subsections we will compare the experiments on these issues and relate them to internal and external validity concerns.

Table 1. Two experiments on observational learning: similarities and differences

	Experiment 1 (Chapter 3)	Experiment 2 (Chapter 4)
Artistic domains	Verbal and visual	Visual
Participants	All students from 6 classes From 3 schools N=131 10th grade Random assignment to conditions	Volunteers 1 School N=61 9th grade Same
Setting	Regular CKV[6] classes, school	Research institute, during free time
Learner characteristics	Initial skill (pre-test score), verbal IQ	Prior knowledge
Comparison group	Practice condition	Practice with direct strategy instruction
Treatment	2 experimental conditions (observational learning with focus on weak model or on strong model)	1 experimental group (observational learning)
Observational learning videos	Strong and weak models (all presented in pairs)	Strong and weak models (sometimes presented in pairs)

[6] *CKV= 'Cultural and Artistic Education', compulsory subject in Dutch secondary education. It includes many artistic domains, such as: visual arts, dance, drama, music, architecture, literature, cinema, etc.*

	Experiment 1 (Chapter 3)	Experiment 2 (Chapter 4)
	Role played scripts Scripts based on spontaneous student behaviour	Fragments from authentic videos Videos were collected when students performed guided tasks
Evaluation task (after watching video)	Identify strong or weak model by comparison (depending on condition), and elaborate: explain choice	Various evaluation and elaboration questions
Product measures	Poems and collages	Design tasks
Assessment procedure	Ratings of creativity Use of anchor products	Ratings of (1) creativity and (2) technique Use of anchor products
Process measures	Indirect measures, on line (poetry), off line (collages)	On line, time sampled self report

1.2.1 Participants

Participants' characteristics differed between the two studies; in the first experiment students from 10th grade participated, while in the second experiment we chose students from 9th grade as we wanted to offer more challenging tasks. In the first experiment, three schools from different parts of the Netherlands participated, while in the second, individual students from just one school were involved. In the first experiment the students participated during their regular classes, while in the second experiment participation was voluntary and during free time.

In both experiments, the students were randomly assigned to conditions to control for many threats to internal validity, such as teacher influence and difference between classes. Internal validity refers to "(…) inferences about whether observed covariation between A and B reflects a causal relationship from A to B in the form in which the variables were manipulated or measured" (Shadish, Cook, & Campbell, 2002, p.53). In other words: internal validity refers to whether the results of an experiment are indeed an effect of the treatment.

External validity refers to generalizability of results over persons, treatments, outcomes and setting (Shadish, et al., 2002, p.83). At first sight, it may seem that the second experiment was more sensitive to external validity threats than the first, because of selective participation. However, we have no specific reasons to think the students in experiment 1 differed from the larger population of 10^{th} grade students: in the first experiment students from three different schools participated and all students in a class participated. In the second experiment, however, possibly only the most motivated students participated: participation was voluntary and the mean intrinsic motivation measured was 5.3 on a 7 point scale (see chapter 4). Moreover, only one school was involved in the second experiment. While the conditions of participation and selection differed for the experiments, similar effects were found. Therefore, we conclude that external validity was not at risk.

1.2.2 Setting

In experiment 1 (chapter 3), the students from both conditions were working in the same classroom. This could easily lead to treatment diffusion, a threat to internal validity. However, since the students worked individually, this threat was not very likely to occur. Moreover, the test tasks were carried out individually and for the visual task a certain physical distance between the students was realized to prevent the students from seeing each other's work. In experiment 2 (chapter 4) the students came to the research institute, which made it easier to provide physical distance between the students and warrant internal validity. We think that in both experiments we optimized circumstances to increase the internal validity, but, as a consequence, we made sacrifices to external validity: the setting was not authentic in either of the experiments. As we explained above, the students worked completely alone, which is not a common condition in Dutch art education. While students in the first experiment worked in their usual classrooms, in the second experiment, the students even came to the research institute. Therefore, we must conclude that, especially for experiment 2, we cannot generalize our findings to normal classroom situations. However, we have shown that the results hold over two different settings (school and research institute) and there are no specific reasons to believe that condition effects would have disappeared by changing the setting.

1.2.3 Learner characteristics

In both experiments we took learner characteristics into account. In the first experiment it was our aim to examine model similarity effects, effects related to the similarity in competence between the observer and the model in the video. Therefore, we included two experimental groups (observation with focus on a weak model and observation with focus on a strong model). We examined whether students with different pre-test score or verbal IQ (in the case of poetry writing) responded differently to these conditions. No model similarity effects were found. Possibly, the observation conditions were too similar as students saw the same videos and only the focus of evaluation of the videos differed.

In experiment 2, only one experimental group was involved (observation). Based on previous research (e.g. Braaksma, Rijlaarsdam, & Van den Bergh, 2002) we assumed that learner characteristics might play a role in the effectiveness of observational learning. Therefore, in experiment 2 we took prior knowledge (the decision to take the visual arts subject in school) into account. We assumed that students who chose the subject visual arts generally possess more drawing skills and more knowledge and experience with the creative process. This appeared to make a difference as prior knowledge interacted significantly with condition regarding certain sub processes and task value. As far as task value is concerned, non-art students generally preferred to watch, perhaps because they think they are lacking drawing skill, while art students preferred to draw. Non-art students sketched less and could be encouraged to sketch by explicit instruction, whereas art students already sketched more at the start of the experiment. They were encouraged to increase sketching time by

observing new sketching strategies. From experiment 2, we received clear indications that including learner characteristics in the research design contributes to instructional theory.

1.2.4 Control group

In the first experiment, we found that task value differed between the conditions at post-test: the students in the observation conditions had higher task value scores than students in the control condition. We wondered whether the positive effect of observation on products and processes could be attributed to this task value difference. Therefore, in the second experiment we aimed at developing an even 'fairer' comparison. We intended to design learning conditions which were equally motivating and, therefore, we provided similar process information to both conditions. So, in experiment 1 we had the students in the control condition practise a task without much process help. In the second experiment, the students in the comparison condition received extensive step wise process guidance based on the model by Sapp (1995). This type of process guidance is not unusual in Dutch art education and it provided process learning for the comparison condition as well.

From the learner reports written after experiment two, we were able to conclude that students in both conditions experienced that they were learning something. While the comparison conditions were different in the two experiments, the results were in favour of the experimental condition(s) for the two experiments in visual art. Such a form of replication contributes to generalizability.

1.2.5 Treatment

In total three treatment variations were used (two observation conditions in the first experiment and one observation condition in the second experiment). In many educational experiments only one treatment variation is studied. Although there are endless possibilities in constructing observational learning videos and although we tested only three, at least several treatment operationalizations were tested, to increase external validity with regard to generalizability of treatments.

1.2.6 Observational learning videos

The videos as well as the supporting lesson materials were constructed differently in the two experiments. While the videos in both experiments were based on authentic task behaviour, the videos in experiment 1 were based on task behaviour without teacher interference or any process guidance, while in the second experiment the videos were authentic videos of students who performed a task guided by process instruction as described by Sapp (1995). Sapp (1995) assumes that the creative process consists of several alternating divergent and convergent processes, starting broadly with several vague idea clusters narrowing down towards one final piece. On the one hand we felt the need for a more structured framework as a basis for the

observational learning materials in experiment two. This would enable us to study the effect on students' processes more closely. On the other hand, we ran the risk of oversimplifying the creative process by reducing complexity and ill-definedness (Efland, 2002). The Sapp model is not a blueprint of a creative process which guarantees a creative result, one may even think that the forced step-by-step process may even hinder creativity. However, very similar process steps are actually used by many teachers. Therefore, we thought this would be a valid process instruction.

While the videos were constructed differently in both experiments, the main difference between the weak and strong model in the videos was similar for both experiments and concerned the difference between a dynamic and a static task approach. In chapter 6 we will elaborate more extensively on this issue as we think this is crucial in constructing observational learning videos about creative processes in art education.

1.2.7 Evaluation task

In the videos we chose to provide contrasting models (strong and weak performers) as often as possible. We thought contrasts would direct the students' attention to the relevant information (Bandura, 1986). We think this was a good decision since reflecting on creative processes is a difficult task. Following Braaksma et al. (2002), we asked students to evaluate the processes shown in the videos by asking them which student of a pair performed better/worse in their opinion. Evaluation as an activity following observation proved to be effective in previous research (Sonnenschein & Whitehurst, 1984). However, in the artistic domain, the issue at stake is whether it is appropriate to ask for qualitative evaluation. Beittel (1972) has shown that there is more than one effective task approach for drawing. We wanted to make students conscious of this variety, avoiding forcing them to use one strategy or the other. Therefore, in experiment 1 we posed the evaluation question in a personal way, asking about the students' own opinion, not about a fixed rule. In the second experiment we varied the questions as well; for instance, we asked the students how they themselves would approach the task or we asked them simply to follow a process more closely (as in Figure 1 in chapter 1). In general, it appeared that the students responded as we intended, as shown by remarks in their learner reports:

> 'I learned that you do not have to do something in one specific way, but you can apply different approaches' (student quote- learner report experiment 2).

1.2.8 Product measures

In the first experiment on observational learning we used a collage task, since Amabile (1982) has shown the effectiveness of collage tasks for measuring creative performance. However, collage making requires a specific visual combinational creativity. A graphic design task as was used in the second experiment may require a different creative skill, which is more related to meaning making and conceptual

creativity. We thought the validity would be enhanced by extending our research to this type of creative performance.

In addition, we added 'technical quality', another product criterion, as a validity check. We did not expect to find a significant difference between the conditions on this criterion, but we examined whether the videos did indeed stimulate creativity rather than the technique. Results supported our expectations: observation promotes creativity, not technique.

One may still question the issue of generalizability, depending on the vision of what art education should be. The tasks the students performed in both experiments: -collages, short poems and graphic design- fit in with a modernist orientation in art education (Emery, 2002), focusing on formal aspects of the work. Moreover, the tasks are not as complex and open as the tasks students have to carry out, for example, in their final examinations at secondary school level. The tasks in our experiments are not ideal tasks for educational ends. They were chosen to provide short and workable tests that fit the experimental conditions we had to deal with. As an experimental setting is a highly controlled environment per se, it was quite difficult to find artistic tasks which would fit such an environment. Tasks related to the postmodern orientation to art education, for instance, generally do not start from one medium or technique. Instead, students are only provided with a vague thematic idea and the teacher coaches the students individually in the meaning making, critical and exploratory process. Decisions about the artistic medium and technique arise from this process, so students are free to choose any artistic medium they think is appropriate in their work. Although we assume that this situation is rather an ideal one and is not encountered in the majority of current art classes, it demonstrates the gap between the tasks we used in the experiments and tasks related to contemporary ideas about art education.

Art education based on a postmodern orientation implies much freedom for the students. However, postmodern tasks require much time and much coaching by the teacher. A student in ninth or tenth grade is not able to perform such a task completely by him/herself and an eight-week task is very common for this type of studio work. Consequently and ideally, products resulting from postmodern art education are very different from each other as students choose their own individual paths. Sometimes, students have the freedom to collaborate in small groups. Therefore, an experimental setting without teacher intervention, a pre-test-treatment-post-test structure, within limited time and resulting in products that can easily be compared and assessed is hard to combine with postmodern art education. Problems would arise with time and assessment in the experimental setting. Therefore, we chose rather short and structured tasks, which are still in use at schools, but not so much in line with theoretical ideas about contemporary art education.

The decision to use short and structured tasks may have caused generalization problems. Implicitly, we assumed that effects we would find for short and structured tasks could be generalized to larger and more complex tasks as in use in contemporary art education. It is questionable, however, if we were allowed do this. Our tasks were relatively open as compared with tasks in structured domains such as math, but

rather structured compared with contemporary art education. As the results of the current experiments suggest no different effect of observation for learning in rather defined and rather ill-defined domains, we may expect similar results for even more ill-defined tasks. More research may shine light on this issue, although, other methods may be used. The methodological concerns we elaborated above may explain the relative absence of experimental studies in contemporary art education research.

1.2.9 Assessment procedure

In the first empirical study of this thesis, the consensual assessment technique (Amabile, 1982) was used for scoring the poems. Amabile (1982) offers an operational definition of creativity and a reliable subjective assessment technique based on this definition: the consensual assessment technique (CAT):

> "A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated. Thus, creativity can be regarded as the quality of products or responses judged to be creative by appropriate observers, and it can also be regarded as the process by which something so judged is produced." (p. 1001)

The CAT is based on an understanding of creativity as a social construct, culture and time specific. Something is considered creative if the social domain considers it to be creative; therefore, creativity should be assessed by people who are experts in the domain. These judges have to rate the products independently and in random order, with their own subjective criteria of creativity and without being trained by the researcher. Scores obtained are no absolute measures of creativity, but relative to one another. The technique holds for ordinary levels of creativity. In the case of extraordinary creativity, the social domain may not immediately recognize the creative potential, which will only be recognized later (for example Vincent van Gogh). According to CAT, other qualities of the work, such as aesthetic appeal and technical quality should be assessed as well in order to check whether creativity can indeed be separated from these qualities. Amabile (1982) demonstrated the reliability of the CAT empirically for the rating of collages and poems.

While this technique worked well for the first empirical study, it requires a considerable number of judges: generally, at least seven judges are used per product. Since many products had to be assessed in the first experiment, we decided to diverge from the CAT. For practical reasons we could only use three judges for each product, therefore we developed a rating scale to support the rating procedure. Products from a previous (consensual assessment) scoring procedures were used as anchor products and judges were asked to compare the products to these anchors. The judges were considered experts in the domain; at least they were students in an art academy (collages) or students in linguistics (poems). They were not trained, but we continued only if they could understand and agree with the anchor products. The judges appeared not to have problems. During the scoring, the judges were familiar with the task of the students. Therefore, it is likely that they used novelty (originality) and appropriateness (to the task) as criteria for judgment, but in fact they were

free to use their own subjective criteria for creative performance. Novelty and appropriateness are often used as criteria for assessing creative products (Amabile, 1982).

As the assessment procedure we used for experiment 1 provided us with reliable scores, it was used in the second experiment as well. However, in the second experiment we asked judges to rate both the creativity and the technical quality of the products in order to check whether we could separate these qualities and to the check the validity of the creativity scores.

1.2.10 Process measures

There are several ways of registering task execution processes: during task execution (on line) and afterwards (off line). Some measures are more direct, getting closer to actual cognitive processes and others are more indirect: studying traces of cognitive activities. Some are based on self reports, while others are more objective measures. Each of these measures has advantages and disadvantages.

In the two experiments creative processes were registered and measured differently. In the first experiment poetry writing processes were registered by using Inputlog, keystroke logging software (Leijten & Van Waes, 2005). This program allowed us to register the writing processes of the students on line (while they were at work) without interfering with these processes. Although we told the students their processes were recorded, they did not notice it. A disadvantage of this approach is that we collected only secondary information about cognitive processes. The physical actions on the key board (keystrokes and pauses) of the students were registered but no information was collected about the mental (cognitive) actions of the students. We had the same problem with the registration of visual creative processes in experiment 1. The students were asked to collect shapes and images cut out for use in their collages but left unused in the end. This measure seemed to be related to revision, as we collected traces of this revision, but no actual cognitive processes were captured.

In the second experiment we aimed at acquiring more detailed process information about the visual creative processes by using an online instrument: time sampled self report (based on Kellogg, 1988; Torrance, Fidalgo, & Garcia, 2007). The students were trained to use it and during their creative processes they marked the activity they were engaged in during each cue. The students were cued to tick a box about every 90 seconds. While this method provided us with detailed process information, in contrast with experiment 1, it may have interfered with the students' processes. Yet, the students did not report that they had experienced any difficulty with regard to this instrument and both conditions in the experiment had to deal with the same circumstances. Therefore, we conclude that it appeared to be a feasible method for registering processes of large groups in experimental settings.

In principle think aloud protocol analyses would provide even more detailed process data, but verbal protocols capture verbal behaviour and in the case of visual tasks, we may wonder whether non verbal behaviour should be included as well or

even the interaction between verbal and physical action (Bar-on, 2007). Besides, this method is not feasible when applied in large classrooms or group experiments. It requires individual sessions and time consuming analyses.

We conclude that we found a useful instrument for process measurement in experiment 2. It combines advantages of both think aloud protocol analyses (online measurement) and questionnaires (feasible for large groups and quickly analysed). However, the underlying assumption is that students are able to determine their own (cognitive) activity. Torrance et al. (2007) have shown that even students in primary education are able to do this for simple physical actions. It may be worthwhile to examine more applications and implementations of this instrument in various contexts. In chapter 6 we will elaborate on creative processes and process measurement.

1.2.11 Artistic domains

Can results be generalized across artistic domains? We found rather different results for poetry writing and visual art production in experiment one. For the visual domain, the creativity of the products was higher for the experimental conditions, while for poetry writing only process results were found. In the verbal domain, the students showed more prewriting and revising after the intervention, but this prewriting and revising did not correlate with the creativity of the products. It appears that new strategies are not directly successful for poetry writing. Braaksma et al. (2002) found positive effects of observation on argumentative writing products and processes. Dutch students have very little experience in poetry writing. Possibly, this explains why they may need more training for products to improve. So, not only domain matters, the task is crucial as well. Still, it seems that process learning is a learning goal in itself. It does not guarantee discovery in poetry writing, though it may increase the likelihood of discovery. According to Beittel (1972) it is not easy to change novice processes. Changing the creative process may be an important learning result itself. Possibly the visual nature of the videos resulted in a more positive effect in the visual domain. In sum, we cannot provide a definite conclusion about the effect of observation in the different artistic domains. We will elaborate on this issue in chapter 6.

1.3 Future research

We have demonstrated the effects of observational learning (experiments in chapter 3 and 4). It seems that observational learning is a powerful learning activity. We would like to put forward some suggestions for further research.

Firstly, after carrying out the experiment we believe that through watching videos and responding to evaluative questions students practise verbalizing and reflecting on creative processes. These skills may be considered important learning outcomes in art education as well. In addition to measuring effects of observation on creative products and processes, other learning outcomes could have been measured

as well, for instance: effects on process knowledge, reflection skills and skills in verbalizing about creative processes.

Secondly, our findings can be extended by more research on how student characteristics influence the effectiveness of observational learning for creative tasks. The experiments we carried out did not provide a definite answer to this question. Does model similarity play a role if the observational learning conditions are more different from each other? The characteristics of the models in observational learning videos may be varied as well; are there different effects if peers or professional artists are models in observational learning videos? For which students would this work or not work?

Thirdly, our experiments took place under highly controlled circumstances and with short and structured tasks. A next step might be to examine the use of observational learning videos in regular classroom situations. As a result, this will require other research methods, for example, ethnographic methods to examine various implementations of observational learning in studio practice. For example, Thomas (2009) observed the teaching of creativity in art education ethnographically. This resulted in a deep understanding of how students learn to be creative in the classroom context.

An option would be to let students produce their own process videos. Choosing important moments or incidents from their own (or other's) processes may be an instructive activity as it would require advanced reflection skills (Leijen, 2008). The production of process videos by students themselves can even replace written process portfolios in upper secondary art education. Constructing process videos may be a form of meaningful reflection for students. Besides, students would develop video editing skills and produce learning materials for each other. We think this is one of many options for the implementation of observational learning in art education and it will be worth studying it.

Fourthly, the effect of observational learning for ill-defined tasks should be studied further. There is a continuum ranging from well defined tasks to extremely ill-defined tasks. Maybe we could study the learning effect of observation for (artistic) tasks with different degrees of ill-definedness as suggested by Van Gog (T. Van Gog, personal communication, April 28, 2009). Does the effect of observation vary or remain constant for these types of tasks? How much time should be spent on training? Do the effects last? How many examples are needed to reach effects? A study on the long term effects of observational learning for ill-defined tasks would supplement our studies.

Besides the suggestion to further examine the effectiveness of observational learning for art education, other interventions which enhance creativity could be studied. In general, we think that we lack knowledge of contemporary instruction methods to enhance creative performance. One of the main goals of secondary art education is to prepare students for higher education in the arts and participation in the contemporary creative economy. Art education is no longer considered only a nice and relaxing activity alternating with 'more important' school subjects. The ultimate goal of art education is to stimulate students' creativity. How can this be

done and reliably measured? To our knowledge, there is not much research. Therefore, we think it is important to conduct more experimental research in art education. What interventions enhance creative processes and products? And what role can ICT play in learning creativity in art education? Attention should be paid to the question of how to measure creative processes and process gain.

It seems that tasks and task perception influence the creative processes considerably (we elaborate on this in chapter 6). Therefore, we think studying the interaction of task and creative process would provide more insight in creative processes. We found that tasks which are relatively structured provide students with 'known solutions'. 'Known solutions' are no new discoveries, but in fact stereotypes, which already exist and which may hinder creative performance. How can tasks provide enough openness and at the same time offer enough structure for students? According to Sapp (1997) it should be possible to adapt a task to the competence level of the student. More advanced students may prefer more open tasks, while less competent students may prefer more structured tasks. Flexible task parameters as suggested by Sapp (1997) may form the basis for an interesting intervention. How can we bring flexible task parameters into practice? Or can we teach students to define a (visual) research question as a start of an unknown discovery process? And how does this affect their creative processes as compared to a 'regular' task?

Beghetto and Kaufman (2007) introduced the idea of mini-c creativity as a central concept for linking creativity and learning. Previous theories on creativity used to distinguish between Big-C and little-c creativity. Big-C creativity is the creativity of eminent persons who create something extraordinary, something completely new in the world. Little-c creativity is a more common form of creativity, everyday creativity. Both Big-C and little-c creativity approaches are rather product oriented; it is about the novelty and appropriateness of the product. Mini-c creativity is about creativity at a detailed level, about, for example, a single decision in a poetry writing process. Mini-c creativity is located within a task or learning process. As Beghetto and Kaufman (2007) suggested, we think that mini-c creativity is a very useful concept. It provides a new, micro level perspective on creativity: what decisions, strategies, approaches appear to be effective? Does divergence, the creation of many alternatives, always work? What role does the-work-in-progress play in the development of the process or at separate mini-c creative moments? More research on mini-c creativity and art education may lead to new insights. We hope that this dissertation will stimulate new research on mini-c creativity in art education.