UNIVERSITY OF TWENTE.

The effect of productive failure instruction on the learning process, learning outcomes, and motivation of young children.

Master thesis

Lotte van Duivenboden
Educational Science and Technology
University of Twente

The effect of productive failure instruction on the learning process, learning outcomes, and motivation of young children.

Master thesis

Examination committee
Prof. dr. A.W. Lazonder
Dr. T.H.S. Eysink

Author
Lotte van Duivenboden
Educational Science and Technology
University of Twente

Date
January, 2017

Acknowledgement

The Master thesis is the final assessment in order to successfully complete the Master programme of Educational Science and Technology at University of Twente. During the thesis I worked with elementary school children on a difficult problem-solving task with LEGO Mindstorms. Where young children were able to adapt quickly to the difficulties, I experienced that failure is sometimes more difficult than I thought. During the process of writing the thesis I have faced some difficulties myself, whereby I have learned to persist even when I failed or got stuck.

This thesis would not have seen the light without help of some people. First of all, I would like to thank my supervisor Ard Lazonder. Thank you for making me come up with my own solutions and willingness to give me feedback when necessary. I would also like to thank my second supervisor Tessa Eysink. Critical comments and suggestions from the both of you have been very much appreciated. Furthermore, my gratitude goes out to the schools and teachers who participated in this research. Especially, my thanks goes out to all the children who participated in this study and worked on a difficult problem-solving task. I enjoyed working with all those children who were really motivated to help me and to play with the LEGO robots. Also many thanks to the Saxion LEGO Education and Innovation studio, they borrowed me the LEGO materials and answered all my questions.

At last, I would like to express my sincere appreciations towards my parents, my friends and my boyfriend. Thank you for listening to all my stories and struggles during the process. I am grateful for their support and advice. They have motivated me to complete the project next to my work as an elementary school teacher.

Lotte van Duivenboden
January, 2017

Abstract

The aim of this study was to design a productive failure lesson which enables elementary school children to become acquainted with programming, using LEGO Mindstorms EV3. This study compared the effectiveness of this lesson on the learning process, the learning outcomes, and motivation of a productive failure lesson with that of a direct instruction lesson. One hundred children from two Dutch elementary schools participated in this study, dyads were composed with regard to the children's ability level. The dyads were assigned to the productive failure condition or the direct instruction condition. During the first session the dyads participated in the lesson, which comprised two tasks, and completed the post-test. The learning process and motivation (before, during and after the tasks) were measured. After a period of two weeks the dyads took the delayed post-test. The results indicated that children were able to learn to program with LEGO Mindstorms EV3. Productive failure yielded comparable learning outcomes and retention as direct instruction. The children in the productive failure condition failed during the first task, on the second task this was less pronounced. A Smileyometer tapped four constructs: self-efficacy, intrinsic motivation, perceived importance of goal mastery and the perceived cause of failure. No significant differences were found regarding intrinsic motivation and the perceived cause of failure. The self-efficacy of the children in the productive failure condition was lower compared to self-efficacy of the children in the direct instruction condition before, during and after task 1. With regard to the perceived cause of goal mastery there were significant differences directly after task 1 and directly after task 2. An implication for future research would be to repeat this study with the same materials but with an improved version of this lesson in which both tasks are combined into one, more difficult, task.

Key words: Productive failure, direct instruction, problem solving, programming.

Table of contents

Introduction

In today's digital society, Dutch citizens are asked to adapt fast to the technological developments. Especially children need to work with the new technologies in the future, therefore it is important to make sure that children become familiar with technology (Rogers & Portsmore, 2004). This notion prompted the development of a new elementary school curriculum in the United Kingdom, which was introduced in 2014. This national curriculum aims to ensure that all children become responsible, competent, creative, and confident users of information and communication technology. One of the core elements of this curriculum is computing education, which includes repeated practice in analyzing problems in computational terms, and solving these problems by writing computer programs (Department for Education, 2013). Thus, learning to program is seen as one way to help children feel more comfortable using technology.

Programming is writing in the language of computers, and Dutch educational boards such as Stichting Kennisnet have argued that, as in the United Kingdom, programming should be taught in elementary schools as a part of science education (Kennisnet, 2015). To teach this skill to children, instruction and opportunities for hands-on practice are needed, just like in 'regular' language learning classes. In most instructional methods, instruction precedes practice. By receiving instruction before practice, children first develop the required knowledge to perform optimal during application exercises. The underlying assumption is that superior performance will automatically lead to superior learning and retention. Unfortunately, this is not necessarily the case, because immediate and delayed performances are not as related as is often believed (Kapur, 2012; Matlen & Klahr, 2012; Schmidt & Bjork, 1992).

As the goal of most instructional methods is to maximize students' performance on the long term (retention) and outside the instructional context (transfer), a solution to reach this goal is called for. An instructional method called *productive failure* could be a fruitful solution. Productive failure starts from the assumption that people can learn from their own mistakes, and that instruction should be delayed until after experimenting (Kapur, 2008). Previous research in high school mathematics lessons showed that productive failure leads to higher scores on a post-test and transfer test than does direct instruction (Kapur, 2012). However, few studies have examined productive failure in an elementary school setting, and the ones that did, produced contrasting findings (DeCaro & Rittle-Johnson, 2012; Fyfe, DeCaro & Rittle-Johnson, 2014). One reason for these contrasting findings could be that children are too young to benefit from an approach that requires them to reflect on their own mistakes (Mazziotti, Loibl & Rummel, 2015). Another reason might be that productive failure involves more than just a change of instructional sequence. According to Kapur and Bielaczyc (2012), 'true' productive failure lessons are based on a set of designated design principles, which will be explained in detail in the section below. Designing for productive failure, in short, involves two phases: a generation phase followed by a consolidation phase. The generation phase enables learners to explore the learning task at hand; the instruction phase that follows consolidates the experiences gained through exploration. Earlier studies of DeCaro and Rittle-Johnson (2012) and Fyfe et al. (2014) may not have not taken the design principles of Kapur and Bielaczyc (2012) into account. The studies changed the sequence of the instruction and evoked the required difficulties during the generation phase, but the instruction during the consolidation phase was not linked to the experiences of the children in the generation phase.

Therefore this research set out to design a productive failure lesson based on the design principles of Kapur and Bielaczyc (2012) using LEGO Mindstorms EV3, which enables upper elementary school children to become acquainted with programming. The effectiveness of this design was evaluated by comparing a productive failure lesson with a direct instruction lesson in terms of learning processes, learning outcomes, and motivation.

Productive failure

Productive failure essentially is the complete opposite of direct instruction. Direct instruction is defined as "providing information that fully explains the concepts and procedures that students are required to learn as well as learning strategy support that is compatible with human cognitive architecture" (Kirschner, Sweller & Clark, 2006, p. 75). In practice this means that a teacher first explains the new concepts step by step and then invites learners to engage in guided practice. In a productive failure lesson, by contrast, the instruction in the new concepts is preceded by unguided exploration. As learners lack the knowledge to successfully solve problems during this exploration phase, they are likely to make mistakes. Although failure seems compelling and is generally considered a negative outcome of the learning process, advocates of the productive failure approach believe that this is not necessarily the case. According to Kapur, Dickson and Yhing (2009) failure can be productive when a situation is created in which children get stuck and in which children receive instruction that is connected to the experiences of the children during the lesson. When the instruction is connected to the experiences of the children, it becomes meaningful and children learn to understand the cause of the failure. Also, VanLehn, Siler, Murray, Yamauchi and Baggett (2003) suggest that successful learning of a new concept is related with failure, which means that it would be better to delay the instruction until after the students reach an impasse and are not able to go on with the task.

To reach such an impasse and to give children the opportunity to learn from their mistakes, it is important that a productive failure lesson is designed according to some guidelines. Kapur and Bielaczyc (2012) made a distinction between two phases; the first phase is a generation phase which is followed by the second phase, a consolidation phase. The goal of the first phase is to generate and explore new ideas and limitations about solution methods. The goal of the second phase is to give children the opportunity to organize and integrate the relevant ideas and solution methods generated in the first phase. The design of both phases involves decisions concerning the activity during the lesson, the participant structures and the social surround. The core design principles of Kapur and Bielaczyc (2012) are stated as follows:

1. Create a problem-solving context that involves working on complex problems that challenge but do not frustrate;
2. Provide opportunities for explanation and elaboration;
3. Provide opportunities to compare and contrast the affordances and constraints with other students.

These design principles indicate that productive failure is not only about changing the sequence of the instruction: the nature of the instruction needs to be designed according to the three core design principles as well. Kapur and Bielaczyc (2012) therefore specified the three core principles into a series of design guidelines (see Table 1). Adhering to these guidelines will result in the design of the two phases under which delaying structure in learning and problem-solving activities may lead to productive failure and effective learning in the long term.

Loibl, Roll and Rummel (2016) also emphasized that a productive failure lesson can only be useful when it is designed according to those design principles. With regard to implementation, three mechanisms are important. First, children should have the prerequisite topical knowledge at the outset of the lesson. Second, children should be confronted with a problem that they have to solve. In a productive failure lesson, children start to look at possible solutions based on the knowledge that they already have. While children are looking for different solutions, they may realize that their knowledge about the topic is too limited. This means that children have to try different solutions based on their prerequisite knowledge and that there is a possibility that they get stuck during the problem-solving task. Third, when children are aware of their knowledge gaps it is important that they recognize deep features of the problem. According to Loibl et al. (2016) this can be realized when, for instance, children get the opportunity to explore and discuss the comparisons and constraints among the self-generated solutions at the beginning of the instruction phase of the lesson (Kapur & Bielaczyc, 2012). When children have activated their prerequisite knowledge, are aware of their knowledge gaps and recognize the deep features of the problem, they eventually receive explicit instruction about the problem-solving task which leads to well-connected and organized knowledge.

Table 1
Core design principles and guidelines for the development of a productive failure lesson.

Phase 1: generation phase		
The activity	Participation structures	Social surround
▪ Find an activity in which students are challenged but not frustrated. ▪ Find problems that require students to make and justify multiple assumptions. ▪ Find an activity that takes the affective draw of the problem scenario into account.	▪ Enable collaboration between children and encourage students to elaborate on the solutions.	▪ Create a safe space to explore and assure students that it is okay to make mistakes or even be unable to solve the problem.
Phase 2: consolidation phase		
The activity	Participation structures	Social surround
▪ Examine student-generated solution methods. ▪ Compare student-generated solution methods with the canonical solution methods.	▪ Enhance engagement through, for instance, group discussions in which students are asked to present the solutions they generated.	▪ Create a safe space to explore the comparisons and constraints among student-generated solutions with the idea that it helps the students to improve their learning.

Note. Based on Kapur & Bielaczyc (2012).

In the studies of Kapur et al. (2009, 2010, 2012) the design principles stated by Kapur and Bielaczyc (2012) were applied to create the mathematics lesson. During the first part of the lesson, the generation phase, students collaborated to explore the problem-solving task and generate new ideas to solve the problem. The teacher did not provide any kind of support or instructional facilitation during this generation phase but assured the students that it was okay to make mistakes or be unable to solve the problem. Students were challenged by the task and they were unable to solve the problems themselves. During the consolidation phase that followed, the teacher started a group discussion and asked the students to share their solutions to the mathematical problems they had to solve. The students looked at comparisons and constraints between the different solutions they generated so that they developed a deep understanding of the concept. Also during this group discussion, the teacher compared the student-generated solutions with the canonical solutions. According to Kapur (2015) "these contrasts afford opportunities to attend to the critical features of the targeted concept that are necessary to develop a deep understanding of the concept" (p. 55). Finally, students were asked to complete well-structured problems to apply the knowledge they had acquired during the group discussion (Kapur et al., 2009, 2010, 2012).

In an earlier study, Kapur (2008) divided a sample of high school students into two groups. The first group received ill-structured problems, which are problems where the goal and ways of reaching it are not easily identified. The second group received well-structured problems, which are problems where all aspects of the problem are clearly specified (Jonassen, 2007). The group that received the ill-structured problems got stuck and failed in most of the cases. Those problems led to more complex, chaotic and divergent discussions when compared to the group that received the well-structured problems. During the problem-solving task the group with the well-structured problems outperformed their ill-structured counterparts on the quality of the solutions. However, the ill-structured group performed better on the post-test, both on well-structured and ill-structured problems. This finding prompted the conclusion that productive failure could be effective. Also, the results of the follow-up studies of Kapur, Dickson and Yhing (2009), Kapur (2010), Schwartz, Chase, Oppezzo & Chin (2011) and Kapur and Bielaczyc (2012) underscore the role of productive failure in learning and show positive results of productive failure instruction on retention and transfer.

However, there are studies which show contradictory results. Fyfe et al. (2014) studied the sequence of instruction and problem solving. The results showed that providing instruction before problem solving leads to retention in contrast to the condition where the instruction was provided after the problem-solving task. A possible reason for the lack of consistent findings regarding the long-term effects of productive failure is that, for instance, the studies of DeCaro and Rittle-Johnson (2012) and Fyfe et al. (2014) did not take the design principles for productive failure into account as stated by Kapur and Bielaczyc (2012). The studies manipulated the sequence of the instruction and evoked the required difficulties during the generation phase, but the instruction during the second phase was not linked to the experiences of the children in the first phase. A second explanation is that most of the studies were conducted in a high school context in which productive failure turned out to be effective (Kapur et al., 2008, 2009, 2010, 2012). It could be that productive failure is not an appropriate instructional method for children in elementary schools. Loibl and Rummel (2014) pointed out that prior knowledge activation is one of the most important learning mechanisms underlying productive failure. Perhaps elementary school children do not have enough prior knowledge to generate different solutions and lack metacognitive and motivational learning strategies to tackle the problem-solving tasks (Mazziotti, Loibl & Rummel, 2015).

In addition, the experience of failure can have several consequences. It could, for example, lead to "negative emotion, lowered self-esteem, reduced intrinsic motivation, and lower expectancies of future success" (Chase, 2011, p.3). In order to learn from failure, children should persist in the task when it becomes difficult. Persistence is included in several motivational theories and various explanations are given for persistence after failure. First, Bandura (1997) pointed out that children who believe in their own capabilities (self-efficacy) are more persistent in completing a task. Second, Deci and Ryan (1985) stated that intrinsic motivation contributes to persistence in a task because children appreciate to work on it. Third, children who decide to master a goal are more persistent because they want to learn to perform the task (Ames & Archer, 1988; Dweck, 1986). Finally, Chase (2011) focused on attribution theory and stated that "the perceived cause of failure is critical" to persist (p.4). According to these results it is reasonable to assume that children's motivation is affected by productive failure instruction, but that they will persist in the learning task if they believe in themselves, are intrinsically motivated, want to master a goal, and know why they have failed.

Research question and hypotheses

In view of the contrasting findings concerning productive failure and the limited amount of research in an elementary school setting, this study set out to establish the effects of productive failure instruction on the learning process, learning outcomes and motivation of upper-elementary school children. During the study, a sample of children from upper-elementary education worked on a complex problem, namely to program a LEGO Mindstorms robot. Children tried to solve this problem in pairs, and each pair was assigned to one of two conditions: the productive failure condition or the direct instruction condition.

The first hypothesis was that children in the productive failure condition would make more mistakes in their LEGO programs during the generation phase of the lesson than children in the direct instruction condition. As children in the productive failure condition received no instruction before they started practice, they were bound to get stuck in the generation phase and, hence, make more mistakes than children in the direct instruction condition.

The second hypothesis stated that the children in the direct instruction condition would have higher scores on the post-test than children in the productive failure condition. This was expected because the children in the direct instruction condition received step-by-step instructions to operate the LEGO Mindstorms equipment, which would enable them to put their newly acquired knowledge in practice. Children in the productive failure lesson received instruction about LEGO Mindstorms *after* practice (i.e., in the consolidation phase) and had to program the robot based on their (insufficient) prior knowledge. This was expected to lead to suboptimal performance during the lesson.

The third hypothesis predicted that children in the productive failure condition would have higher scores on the retention test than children in the direct instruction condition. Previous research showed positive results of productive failure instruction on retention and transfer (Kapur, Dickson & Yhing, 2009; Kapur, 2010; Schwartz et al., 2011). The productive failure lessons in these studies were based on the design principles and guidelines summarized in Table 1. As the present study used the same design principles, positive effects on retention were expected to show, regardless of the fact that the sample comprised a younger group of learners.

The fourth hypothesis was that the motivation of the children in the productive failure condition was not equal to the motivation of the children in the direct instruction condition. Because children in the productive failure condition were expected to get stuck and fail on the problem-solving task, it was the question what this would do to their motivation. Two possible scenarios exist, one negative and one positive. When children get stuck during the problem-solving task, this could have a detrimental effect on motivation, causing children to become frustrated and give up trying. Another possibility might be that children, in particular the more persistent ones, were positively challenged by the difficult task they were facing, and became determined and motivated to solve it. This, in turn, would increase their performance during the generation phase, as well as their receptiveness to learning the canonical solution in the consolidation phase.

Method

Research design

This study used an experimental between-group design to evaluate the effectiveness of the productive failure method. The independent variable in this study was the instructional method; dependent variables relate to learning processes, learning outcomes, and motivation. Participants were randomly assigned to either the experimental condition (productive failure) or the control condition (direct instruction). Both groups were given an immediate and a delayed post-test to assess learning outcomes. Motivation was measured before, during, and after the instructional intervention.

Participants

A total of 100 children without prior experience with LEGO Mindstorms participated in this study. As children were between 10 and 12 years old, parental permission was needed. The school principal informed the parents and care-takers about the study and gave them the opportunity to object against their child's participation—which none of them did. The children (60 boys and 40 girls) came from two schools in the Netherlands and were between 10 and 12 years old (M = 11.18, SD = 0.74). As Kapur and Bielaczyc (2012) emphasized the importance of collaboration, children in each class were grouped in dyads who were then assigned to conditions. First, dyads were composed based on children's scholastic achievement in comprehensive reading (i.e., CITO scores) because programming could be compared to this subject (Kennisnet, 2015). The sampling strategy was heterogeneous sampling: high-performing children collaborated with average-performing children and low-performing children also collaborated with average-performing children . Second, the 50 dyads were randomly and equally assigned to either the productive failure condition or the direct instruction condition.

Learning task and instructional materials

During the lesson, all children were acquainted with programming using LEGO Mindstorms EV3, a programmable robotics construction set that enables children to build, program and command their own LEGO robots (see Figure 1). To measure the effects of the instructional interventions, one lesson was designed, that differed regarding the instructional method (i.e., productive failure or direct instruction; see Table 2). The lesson comprised two tasks (see Table 3) in which children eventually learned to program their LEGO Mindstorms EV3 robot car to drive a specific route and to park their car in a parking lot.

Figure 1. *LEGO Mindstorms EV3 car.*

Table 2
Productive failure instruction and direct instruction.

	Productive failure	Direct instruction
Phase 1	*Exploration* Exploring the problem-solving task in dyads without support or instructional facilitation of the researcher.	*Instruction* The researcher presents a quick activity that engages children's thinking, states the objective of the lesson, provides the rationale of the lesson and presents the content of the lesson and gives a demonstration of the skills and procedures. The researcher concludes with asking questions to assess children's understanding and provides guided practice.
Phase 2	*Instruction* The researcher asks the dyads to share their solution to the learning task. The children compare their solution to the solution of another dyad. The researcher compares the children's solution to the canonical solution and gives instruction about the task and the canonical solution. Finally, the children practice with a different well-structured task.	*Exploration* Finally, the researcher reviews the lesson and points out the importance of the content.

Note. In both conditions, each task of the lesson consisted of two phases; an exploration phase and an instruction phase. In the productive failure condition the exploration phase is equal to the generation phase and the instruction phase is equal to the consolidation phase according to Kapur and Bielaczyc (2012).

Table 3
Content of the lesson.

	Content	Goal	Example
Task 1	Using the action block 'big motor'.	The children learn to program the car so that it could change directions.	The children are able to program the car so that it could change directions (right and left).
Task 2	Using the action block 'changing directions' and combining both action blocks.	The children learn to program the car so that it drives straight ahead. Also, children learn to combine action blocks.	The children are able to program the car so that it drives straight ahead, drives to the right and drives to the left.
Post-test	Combining both action blocks.	The children learn to combine the different blocks and are able to drive a route with the LEGO car.	The children are able to program the car to drive a designated route and to park the car.

To program the LEGO Mindstorms EV3 car, corresponding software was used. The children used action blocks and variables in the LEGO Mindstorms software (see Figure 2). The green action blocks control the actions of the program. For instance, in Figure 2, the LEGO Mindstorms car drives straight ahead, turns to the right and then turns to the left. Variables are part of the action blocks. For instance, in Figure 2, within the first action block there are different variables such as speed, motor rotations and changing directions. The programs that the children wrote were saved in the LEGO Mindstorms EV3 software.

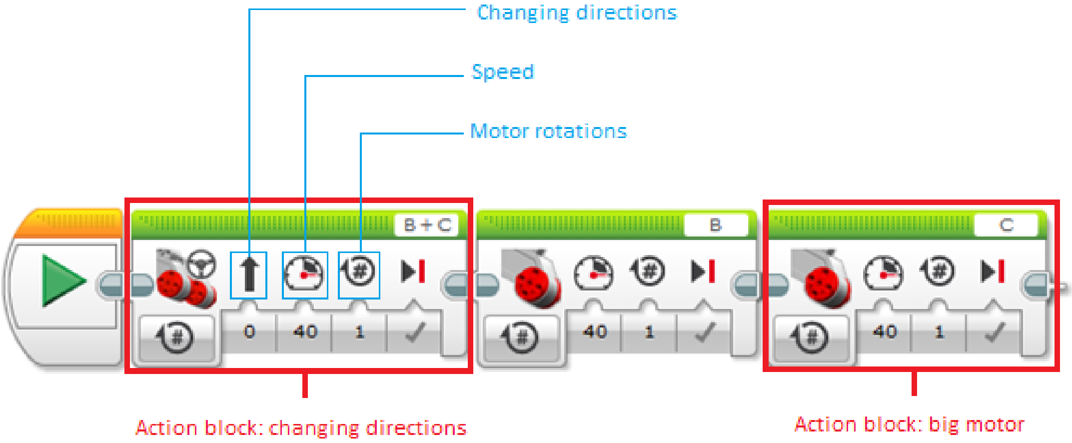

Figure 2. *Program for the LEGO Mindstorms EV3 car.*

Instruments

Immediate and delayed post-test

Learning outcomes were assessed immediately after the instructional intervention with an immediate post-test. After two weeks, the children received a delayed post-test to measure retention, which was identical to the immediate post-test. Both tests asked the dyads to program the LEGO Mindstorms robot such that it drives a designated route and stops in a parking lot. This assignment was practiced during the productive failure instruction and the direct instruction.

Smileyometer

The Smileyometer (Read, 2007) is an instrument from the Fun Toolkit, an instrument that can be used to gather opinions from children about technological applications, in particular their expectations, experienced feelings or fun. The Smileyometer contained pictorial representations (see Figure 3) that formed a 5-point Likert scale. In this study the Smileyometer was used to measure the motivation of the children before, during, and after the instructional intervention. The first question, “I am smart enough to make the LEGO car move”, measured the self-efficacy of the children. The second question, “I like learning to make the LEGO car, measured intrinsic motivation”. The third question, “I think it is important to know how to make the LEGO car move”, measured the degree to which children wanted to master the goal. The last question, “I think it is important to put in more effort when I have made a mistake with the LEGO car”, measured if the children attributed the cause of failure to themselves.

Figure 3. *Smileyometer.*

Procedure

Every dyad attended two sessions, which were guided by the researcher. The researcher received two dyads at a time, so that the dyads in the productive failure condition were able to discuss their findings after the generation phase. During the first session, the productive failure and direct instruction lesson were delivered by the researcher, who also administered the Smileyometer and immediate post-test. The delayed post-test was administered during the second session, which took place two weeks later. The first session started with an introduction about the aim of the study and the researcher introduced the Smileyometer and the learning task. After this general introduction, children in both conditions filled out the Smileyometer to assess their motivation before the task.

The lessons in both conditions were built around the two tasks described in Table 3. In the productive failure condition the children explored the first problem-solving sub-task in dyads. The researcher did not give support or instructional facilitation. After the generation phase, the children received the Smileyometer for the second time. Next, the researcher asked the dyads to share their solution to the problem-solving task. The researcher compared the children's solution to the canonical solution. The researcher provided the children with a well-structured task so that they could apply the knowledge they obtained during the lesson. This procedure was repeated for the second task.

In the direct instruction condition, the researcher started with instruction. She presented a quick activity that engages children's thinking, stated the objective of the lesson, provided the rationale of the lesson, presented the content of the lesson, and demonstrated the skills and procedures needed to program the LEGO Mindstorms car. The researcher concluded by asking questions to assess children's understanding. After the instruction, the children received the Smileyometer for the second time. After filling in the Smileyometer, the researcher provided guided practice and when the children were proficient enough, they could practice independently. The researcher reviewed the lesson and pointed out the importance of the content. Also in the direct instruction, the researcher provided the children with a well-structured task. This procedure was repeated for the second task.

After the interventions, children from both conditions received a Smileyometer for the third time. They were then asked to complete the immediate post-test. The program that the children wrote was saved into the LEGO Mindstorms EV3 software and the quality of the program was analysed afterwards. The second session followed after two weeks. During the second session the children received the delayed post-test.

Coding and scoring of the dependent variables

Children's learning process was measured during the two tasks. The goal was to examine if failure took place during the productive failure lesson and the direct instruction lesson. To measure the learning process, the dyads received 1 point when they were able to complete the task. Dyads who were unable to complete the task received 0 points. Second, the number of attempts per task was scored. Third, the dyads in both conditions got scores for each action block as well as for each variable that they have programmed in a correct way. Because there were multiple solutions that could be used and not all the solutions yielded the same number of points, scores were converted to percentages. The attempt with the best solution was used for analysis of the learning process. To assess the inter-rater reliability of the learning process, a second coder rated 50 programs (50% of total programs). The inter-rater reliability coefficient was measured with a k = .81 (Cohen's Kappa).

The programs that the dyads designed were scored similar to the learning process. Also the designed programs were scored on completion, correct action blocks and correct variables. The number of attempts was not scored because the children received one chance during the post-test and the delayed post-test. To assess the inter-rater reliability for the learning outcomes, a second coder rated 25 post-tests (50% of total post-tests). The inter-rater reliability coefficient was measured with a k = .89 (Cohen's Kappa).

Results

Learning process

Task 1

A chi-square test of independence was used to assess whether the number of dyads who successfully completed the programming task differed across conditions. The frequency counts showed that 11 of the 25 dyads in the productive failure condition successfully completed the first programming task during the exploration phase. All 25 dyads in the direct instruction condition completed the first task successfully. This difference was significant, $\chi^2(1) = 19.44$, $p < .001$, and odds ratio revealed that dyads who received direct instruction were 64.30 times more likely to successfully complete the first programming task than children in the productive failure condition.

Table 4

Means and standard deviations of task 1.

	Productive failure (n = 25)		Direct instruction (n = 25)	
	M	*SD*	*M*	*SD*
Exploration				
Number of attempts	3.96	1.43	1.36	0.49
Mistakes in action blocks (%)	28.67	32.10	0.00	0.00
Mistakes in variables (%)	23.68	13.13	0.00	0.00
Instruction				
Number of attempts	1.04	0.20	1.00	0.00
Mistakes in action blocks (%)	0.00	0.00	0.00	0.00
Mistakes in variables (%)	0.00	0.00	0.00	0.00

Note. The exploration phase is for the productive failure condition the first phase of the task, the instruction phase is the second phase of the task. For the direct instruction condition, the exploration phase is the second phase of the task and the instruction phase is the first phase of the task.

The programs children generated during the exploration phase were analyzed for the number of attempts and mistakes. The data in Table 4 show that dyads in the productive failure condition made nearly four attempts on average to try to build a correct program whereas most dyads in the direct instruction condition needed a single attempt. Preliminary checks revealed non-normality in the scores during the first, exploration, phase on the action blocks, $D(50) = 0.44$, $p < .001$, and variables, $D(50) = 0.34$, $p < .001$. Therefore, Mann-Whitney U tests were used to assess whether the number of attempts and mistakes differed between conditions. The results showed that the number of attempts in the productive failure condition (Mdn = 4.00) differed significantly from that in the direct instruction condition (Mdn = 1.00), $U = 18.00$, $z = -5.88$, $p < .001$, $r = -.83$. Table 4 also shows the mistakes in the action blocks and variables. The mistakes in the action blocks in the productive failure condition (Mdn = 33.33) differed significantly from that in the direct instruction condition (Mdn = 0.00), $U = 150.00$, $z = -4.10$, $p < .001$, $r = -.58$. Also, on the variables there was a significant difference between the productive failure condition (Mdn = 21.43) and the direct instruction condition (Mdn = 0.00), $U = 37.50$, $z = -5.88$, $p < .001$, $r = -.83$.

The results of the exploration phase of the productive failure condition were compared to the results of the instruction phase to assess whether dyads improved their performance after instruction. Using Wilcoxon signed-rank tests, the number of attempts were found to be significantly higher in the exploration phase (*Mdn* = 4.00) than in the instruction phase (*Mdn* = 1.00), $z = -4.397$, $p < .001$, $r = -.62$. The mistakes in the action blocks were also significantly higher in the exploration phase (*Mdn* = 33.33) than in the instruction phase (*Mdn* = 0.00), $z = -3.241$, $p = .001$, $r = -.46$. Also on the variables, the mistakes were significantly higher in the exploration phase (*Mdn* = 21.43) than in the instruction phase (*Mdn* = 0.00), $z = -4.119$, $p < .001$, $r = -.58$.

Task 2

As in the previous task, scores during the exploration phase on the action blocks, $D(50) = 0.54$, $p < .001$, and variables, $D(50) = 0.47$, $p < .001$, were both significantly non-normal and so were the scores on the completion of the second task, $D(50) = 0.54$, $p < .001$. Therefore, the same non-parametric tests were used to analyse the data. Frequency counts showed that 23 of the 25 dyads in the productive failure condition successfully completed the programming task during the exploration phase of the second task. In the direct instruction condition all dyads successfully completed the task in the exploration phase. The number of dyads who completed the task was independent of condition, $\chi^2(1) = 2.08$, $p = .149$

Table 5

Means and standard deviations of task 2.

	Productive failure (n = 25)		Direct instruction (n = 25)	
	M	*SD*	*M*	*SD*
Exploration				
Number of attempts	1.88	0.88	1.00	0.00
Mistakes in action blocks (%)	4.00	14.66	0.00	0.00
Mistakes in variables (%)	4.86	8.99	0.31	1.54
Instruction				
Number of attempts	1.08	0.40	1.00	0.00
Mistakes in action blocks (%)	0.00	0.00	0.00	0.00
Mistakes in variables (%)	0.53	2.67	0.00	0.00

Note. The exploration phase is for the productive failure condition the first phase of the task, the instruction phase is the second phase of the task. For the direct instruction condition, the exploration phase is the second phase of the task and the instruction phase is the first phase of the task.

The data of Table 5 show that the number of attempts in the productive failure condition (*Mdn* = 2.00) differed significantly from that in the direct instruction condition (*Mdn* = 1.00) during the exploration phase of the second task, $U = 112.50$, $z = -4.73$, $p < .001$, $r = -.67$. The mistakes in the action blocks in the productive failure condition (*Mdn* = 0.00) did not differ significantly from the mistakes in the action blocks in the direct instruction condition (*Mdn* = 0.00), $U = 287.50$, $z = -1.43$, $p = .153$, $r = -.20$. The mistakes in the variables differed significantly between the productive failure condition (*Mdn* = 0.00) and the direct instruction condition (*Mdn* = 0.00), $U = 222.50$, $z = -2.61$, $p = .009$, $r = -.37$. All dyads were able to complete the programming task successfully after the instruction in the productive failure condition as well as in the direct instruction condition.

The results of the exploration phase of the productive failure condition were compared to the results of the instruction phase to assess whether dyads improved their performance after instruction by using Wilcoxon signed-rank tests. The number of attempts in the productive failure condition was significantly higher in the exploration phase (*Mdn* = 2.00) than in the instruction phase (*Mdn* = 1.00), $z = -3.345$, $p = .001$, $r = -.47$. The mistakes in the action blocks in the productive failure condition did not differ significantly between the exploration phase (*Mdn* = 0.00) and the instruction phase (*Mdn* = 0.00), z = -1.342, p = 0.18, r = -.19. However, the mistakes in the variables was significantly higher in the exploration phase (*Mdn* = 0.00) than in the instruction phase (*Mdn* = 0.00), z = -2.111 , p = .035 , r = -.30.

Learning outcomes

Children's performance on the immediate post-test pointed to a ceiling effect (see Table 6) and preliminary checks confirmed non-normality. The post-test scores on the action blocks, $D(50) = 0.51$, $p < .001$, and variables, $D(50) = 0.39$, $p < .001$, were both significantly non-normal. Therefore a chi-square test of independence was used to analyse whether the number of dyads who successfully completed the post-test differed across the direct instruction condition and the productive failure condition. The frequency counts showed that 18 of the 25 dyads in the productive failure condition and 21 of the 25 dyads in the direct instruction condition made a correct program on the post-test. This success ratio was independent of condition, $\chi^2(1) = 1.05$, $p = .306$.

Mann-Whitney U tests were conducted to assess whether the mistakes differed between the productive failure and the direct instruction condition. The mistakes in the action blocks in the productive failure condition (*Mdn* = 0.00) did not differ significantly from the mistakes made in the direct instruction condition (*Mdn* = 0.00), U = 285.50, z = -0.93, p = .353, r = -.13. Also on the variables there was no significant difference between the productive failure condition (*Mdn* = 0.00) and the direct instruction condition (*Mdn* = 0.00), U = 247.50, z = -1.64, p = .102, r = .23.

Table 6

Means and standard deviations of the post-test and the retention test.

	Productive failure (n = 25)		Direct instruction (n = 25)	
	M	*SD*	*M*	*SD*
Post-test				
Score for action blocks (%)	94.48	14.33	98.19	6.49
Score for variables (%)	96.08	11.27	99.39	1.54
Retention test				
Score for action blocks (%)	93.05	14.42	94.86	12.96
Score for variables (%)	97.10	3.72	97.84	4.00

The retention test assessed whether and to what extent children managed to maintain their performance level over a longer period of time. These results too pointed to a ceiling effect. The retention test scores on the action blocks $D(50) = 0.47$, $p < .001$ and the retention test scores on the variables, $D(50) = 0.35$, $p < .001$, were both significantly non-normal. Therefore, a comparison was made between the dyads who successfully completed the post-test as well as the retention test. The frequency counts show that 17 out of the 21 dyads in the direct instruction condition were successful on both tests. In the productive failure condition, 11 out of the 18 dyads successfully completed the post-test as well as the retention test. The number of dyads who successfully completed the retention test was independent of condition, $\chi^2(1) = 0.40$, $p = .529$.

Mann-Whitney U tests were conducted to assess whether the mistakes differed between the productive failure and the direct instruction condition over a longer period of time. The mistakes on the action blocks in the productive failure condition (*Mdn* = 0.00) did not differ significantly from the mistakes made in the direct instruction condition (*Mdn* = 0.00), U = 288.50, z = -0.67, p = .505 , r = -.09.

Also on the variables there was no significant difference between the productive failure condition (*Mdn* = 0.00) and the direct instruction condition (*Mdn* = 0.00), $U = 261.00$, $z = -1.13$, $p = .258$, $r = -.16$.

Motivation

Mann-Whitney U tests were conducted to determine whether motivation differed between the productive failure condition and the direct instruction condition. The results are presented in Table 7.

Table 7
Mann Whitney U test scores regarding the Smileyometer.

	Productive failure (n = 50)	Direct instruction (n = 50)	U	z	p	r
	Median	*Median*				
Self-efficacy						
Before Task 1	3.00	4.00	1638.00	2.90	.004	.29
During Task 1	3.00	4.00	1811.00	4.10	<.001	.41
Before Task 2	4.00	5.00	1695.00	3.35	.001	.34
During Task 2	4.00	5.00	1502.00	1.94	.053	.19
After Task 2	5.00	5.00	1464.50	1.73	.084	.17
Intrinsic motivation						
Before Task 1	5.00	4.00	1156.00	0.73	.466	.07
During Task 1	5.00	5.00	1321.00	0.56	.577	.06
Before Task 2	5.00	5.00	1389.00	1.13	.257	.11
During Task 2	5.00	5.00	1257.50	0.06	.948	.01
After Task 2	5.00	5.00	1263.00	0.12	.906	.01
Goal mastery						
Before Task 1	3.00	3.00	1395.00	1.08	.282	.11
During Task 1	3.00	4.00	1434.00	1.36	.173	.14
Before Task 2	3.00	4.00	1539.50	2.11	.035	.21
During Task 2	4.00	4.00	1440.00	1.39	.165	.14
After Task 2	4.00	4.00	1546.00	2.16	.029	.22
Attribution						
Before Task 1	4.00	4.00	1404.00	1.16	.246	.12
During Task 1	4.00	4.00	1368.50	0.87	.386	.09
Before Task 2	4.00	4.00	1396.50	1.09	.278	.11
During Task 2	4.00	4.00	1251.00	0.01	.994	.00
After Task 2	4.00	4.00	1354.00	0.78	.437	.08

The self-efficacy of the children in the productive failure condition differed significantly from the children in the direct instruction condition before the first task. Also during the first task the self-efficacy of the children in the productive failure condition differed significantly from the children in the direct instruction. And after the first task the self-efficacy of the children in the productive failure condition differed significantly from the children in the direct instruction. There were no significant differences between the productive failure condition and the direct instruction condition regarding self-efficacy during task 2 and after task 2.

With regard to intrinsic motivation, there were no significant differences between the productive failure condition and the direct instruction condition.

The perceived importance of goal mastery of the children in the productive failure condition differed significantly from the children in the direct instruction condition after the first task. Also after the second task the perceived importance of goal mastery of the children in the productive failure condition differed significantly from the children in the direct instruction condition. There were no significant differences between the productive failure condition and the direct instruction condition regarding the perceived importance of goal mastery before task 1, during task 1 and during task 2.

With regard to the attribution of the perceived cause of failure, there were no significant differences between the productive failure condition and the direct instruction condition.

Conclusion and discussion

The aim of this study was to design a productive failure lesson which enables elementary school children to become acquainted with programming, using LEGO Mindstorms EV3. This study compared the effectiveness of this lesson on the learning process, the learning outcomes and motivation of a productive failure lesson with that of a direct instruction lesson.

Because the children in the productive failure condition were bound to get stuck, it was expected that children in the productive failure condition would make more mistakes than children in the direct instruction condition. Results regarding the learning process are partly consistent with this hypothesis. Children who received productive failure instruction indeed made more mistakes during the exploration phase of the first task than the children who received direct instruction. Also, the number of attempts in the productive failure condition was higher than in the direct instruction condition, which points to failure in the productive failure condition because children needed to try different solutions and were not able to complete the task in a single attempt. During the second task the differences between the conditions with regard to the attempts was less pronounced. As in the first task, children in the productive failure condition needed to do more attempts to complete the programming task. There were no significant differences during the second task in the exploration phase regarding the number of mistakes.

With regard to learning outcomes, it was expected that children in the direct instruction condition would outperform their counterparts from the productive failure condition directly after the lesson. The results of the post-test do not support this hypothesis: children in both conditions performed as well on the post-test. There were no significant differences between both conditions with regard to the learning outcomes. This result is in contradiction of the results of Kapur (2008, 2010, 2011, 2012), where children in the direct instruction condition performed better on the immediate post-test than the children in the productive failure condition. As mentioned before, the children failed during the first task, however during the second task failure did not take place in the productive failure condition. As a consequence, it is not possible to conclude that productive failure instruction is equally effective as direct instruction. If that were the case, there should have been failure in both tasks. The third hypothesis stated that children in the productive failure condition would have higher scores on the retention test than children in the direct instruction condition. There were no significant differences between both conditions on this measure, meaning that the children in the direct instruction condition performed better than expected and retention took place in both conditions. This result too contradicts the outcomes of previous studies (Kapur, 2008, 2010, 2011, 2012).

The results with regard to the learning process and the learning outcomes could be questioned. The results are partly in line with the hypotheses, but showed some surprising results. As mentioned before the children did fail on the first task, but did not on the second task. This means that the conditions for productive failure were not met on this task. But why did children not fail on the second task? Probably because the second task was too easy for them. Children practiced with the materials during task 1 and improved their performance with regard to programming, which enabled them to use the knowledge they obtained during task 1 for task 2. A second explanation is that both tasks were rather limited in scope. It was expected that learning to program would be difficult for children without prior knowledge during their first encounter with the LEGO Mindstorms EV3 materials, therefore it was decided to keep the scope of the tasks small. It could be that both tasks were too similar whereby the children could complete the second task on the basis of the instruction of the first task. An implication for future research would be to repeat this study with the same materials but with an improved version of this lesson in which both tasks are combined into one, more difficult, task. It would be interesting to focus again on the learning process, learning outcomes and motivation when the children fail on the task. Only then it is possible to get a distinct view of whether productive failure actually is productive.

The fourth hypothesis stated that the motivation of the children in the productive failure condition was not equal to the motivation of the children in the direct instruction condition. Two possible explanations were given. The first predicted a decrease in motivation because children could become frustrated and even give up trying. The second explanation, in contrast, predicted an increase in motivation because children, in particular the more persistent ones, would be positively challenged by a difficult task and would become determined and motivated to solve it. The motivation of the children was measured by a Smileyometer that tapped four constructs: self-efficacy, intrinsic motivation, perceived importance of goal mastery and the perceived cause of failure. No significant differences were found regarding intrinsic motivation and the perceived cause of failure. Notable is that the self-efficacy of the children in the productive failure condition was lower than the self-efficacy of the children in the direct instruction condition only before, during and after task 1. An explanation with regard to the differences in self-efficacy before task 1 is that the children in the productive failure condition already knew that they had to explore the task by themselves. The children in the direct instruction condition knew that they would have instruction before exploration. An explanation for the differences during and after task 1 is that most dyads in the productive failure condition failed on the first task, they needed to try multiple solutions or they were unable to complete the task. The experience of failure during task 1 could lead to the observed differences in self-efficacy during and after task 1. There were no differences regarding self-efficacy during and after task 2. With regard to the perceived cause of goal mastery there were significant differences directly after task 1 and directly after task 2. The children in the direct instruction condition seem to be more convinced that goal mastery is important than the children in the productive failure condition. A possible explanation is that children in the direct instruction condition knew what the goal of the task was, this was mentioned at the beginning of the instruction and the instructor demonstrated the task, which showed what the children eventually were expected to do. This was not the case in the productive failure condition, where children received the task and were asked to explore it by themselves. In some cases the children were that enthusiastic about the LEGO robots that they forgot what the goal of the task was.

Future research is needed to examine to what extent productive failure is an effective method for early programming instruction. The instructor worked with two dyads at a time in a separate room. In future research it is interesting to apply productive failure instruction in regular classrooms. In practice teachers mostly have a class with approximately 30 children at the same time. Teachers are not always able to give small-group instruction, therefore it would be recommended to focus on productive failure instruction in a more authentic context for the teachers. Second, this study took a first step in investigating the effect of productive failure instruction on motivation. It is recommended to dig deeper into the field of motivation in combination with productive failure instruction. It seemed that children liked the productive failure lesson and were enthusiastic despite the experience of failure during task 1. Especially the children in the productive failure condition seemed proud of themselves at the end of the lesson, probably because they realized that they were able to complete the difficult task on their own. Because children did not fail during the second task it is important to look again at the motivational and affective aspect with regard to productive failure in combination with a more difficult task.

As a result of this study there are some practical implications for elementary education. First, this study showed that children are able to learn to program with LEGO Mindstorms EV3. Where teachers in elementary schools often expect that programming is a skill for the high-performing children, this study showed that all children are able to program a LEGO car to drive a designated route. This means that in elementary education programming does not have to be only for the high-performing children. Second, productive failure seems to be a method which is appropriate for science and technology education. An important principle of productive failure instruction is that children should work on problems that challenge but do not frustrate (Kapur & Bielaczyc, 2012), in practice this means that children should get stuck and make mistakes. Those moments of impasse give children the opportunity to realize that their way of thinking may be incorrect. After these moments of insight, children could explore other possible solutions. Finally, the results of this study showed that children in both conditions were able to learn how to program the LEGO Mindstorms EV3 robot. It seems that it makes no difference which instructional method is used with regard to the learning outcomes. The productive failure instruction did not impede children's learning -- but did not have added value with regard to learning outcomes either. In practice this means that teachers could decide which instructional method fits the content of the lesson best and if children should explore before instruction or the other way around.

References

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Chase, C. (2011). *Motivating persistence in the face of failure: the impact of an ego-protective buffer on learning choices and outcomes in a computer-based educational game* (Dissertation). Retrieved from: https://stacks.stanford.edu/file/druid:yw409rc6957/Chase_Dissertation_Submitted_to_Stanford_FULL-augmented.pdf

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*, 552-568.

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

Department for Education (2013). National curriculum in England: computing programmes of study. Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study/national-curriculum-in-england-computing-programmes-of-study

Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology, 84*, 502-519.

Jonassen, D. (2007). Engaging and supporting problem solving online. In S. M. Fiore & E. Salas (Eds.), *Toward a science of distributed learning* (pp. 185-207). Washington, DC: American Psychological Association.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*, 379-424.

Kapur, M., Dickson, L., & Yhing, T.P. (2009). Productive failure in mathematical problem solving. *Instructional Science, 38,* 1717-1722.

Kapur, M. (2010). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39*, 561–579.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*, 651-672.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*, 289–299.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21,* 45-83.

Kennisnet (2015a). *Alles wat je moet weten over programmeren en maken.* Retrieved from: https://www.kennisnet.nl/artikel/alles-wat-je-moet-weten-over-programmeren-en-maken/

Kennisnet (2015b). *Wat wij kunnen leren van computing-onderwijs in de Britse praktijk.* Retrieved from: https://www.kennisnet.nl/artikel/wat-wij-kunnen-leren-van-computing-onderwijs-in-de-britse-praktijk/

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educational Psychologist, 41,* 75–86.

Loibl, K., Roll, I., & Rummel, N. (2016). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review,* 1-23. doi: 10.1007/s10648-016-9379-x.

Loibl, K. & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and*

Instruction, 34, 74-85.
Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*, 621-634
Mazziotti, C., Loibl, K., & Rummel, N. (2015). *Collaborative or individual learning within productive failure; Does the social form of learning make a difference?* CSCL 2015, Gothenburg, Sweden.
Read, J.C. (2007). Validating the Fun Toolkit: An instrument for measuring children's opinions of technology. *Cognition, Technology & Work, 10,* 119-128.
Rogers, C. & Portsmore, M. (2004). Bringing engineering to elementary school. *Journal of STEM Education, 5,* 17-28.
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.
Schwartz, D.L., Chase, C.C., Oppezzo, M.A., & Chin, D.B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103,* 759-775.
VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21,* 209– 249.