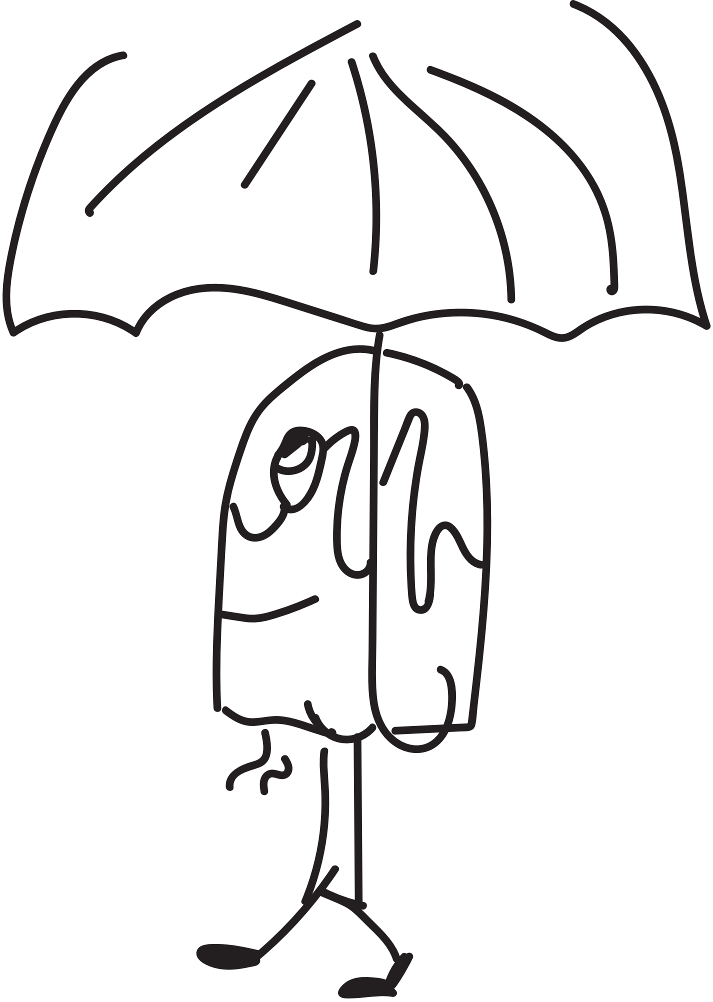


ASSESSING CREATIVE PROBLEM SOLVING IN PRIMARY SCHOOL STUDENTS

Mare van Hooijdonk

Dissertation

ASSESSING CREATIVE PROBLEM SOLVING IN PRIMARY SCHOOL STUDENTS

Mare van Hooijdonk

Dissertation

ISBN: 978-94-93315-54-9

DOI: https://doi.org/10.33540/1726

Printed by: proefschrift-aio.nl

Design and Layout: Marie Cecile Oosterhout - studio-bont.nl

Illustrations: The children that participated in the studies.

ASSESSING CREATIVE PROBLEM SOLVING IN PRIMARY SCHOOL STUDENTS

Creatief Probleemoplossen Meetbaar Maken in het Basisonderwijs

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 26 mei 2023 des middags te 2.15 uur

door

Mare van Hooijdonk

geboren op 3 mei 1988
te Nijmegen

Promotors:

Prof. dr. E. H. Kroesbergen
Prof. dr. M. T. Mainhard
Prof. dr. J. van Tartwijk

Beoordelingscommissie:

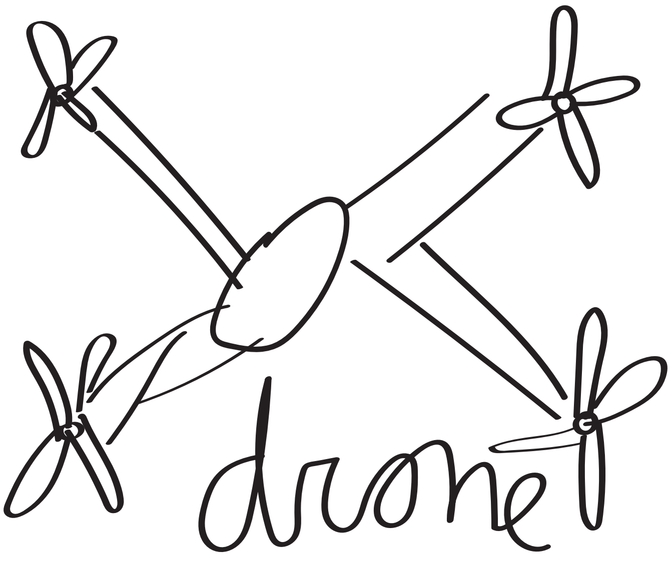
Prof. dr. M. Karwowski
Prof. dr. L. Kester
Prof. dr. P.P.M. Leseman
Prof. dr. P.C. Meijer
Prof. dr. M.E.J. Raijmakers

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van het Ministerie van Onderwijs, Cultuur, en Wetenschap (OCW/PromoDoc/1065001).

CONTENTS

1

GENERAL INTRODUCTION



"Maybe um. Do you have one at home? A drone. And then, there's a container under it and you can put the ice pops in there. And then it brings it home. And then your parents put it in the freezer. Or something like that?"

The idea presented on the previous page was generated by a ten-year-old fourth grader as part of a task. In this task, she was asked to come up with ideas to help Lisa and Simon to bring ice pops home from the supermarket on a hot day. Of course, the challenge is to avoid that the ice pops melt. It is not difficult to imagine that the creative ideas of this primary schooler could be (and actually are) translated into innovations applied in practice.

Carrying out this task successfully is an example of creative problem-solving (CPS), a form of creativity in which creative abilities and knowledge are used to solve everyday problems (Isaksen et al., 2011). CPS is regarded as an essential ability to deal with the uncertainties and possibilities that modern society brings (Craft, 2011). When CPS was applied in the primary school context, it not only showed to enhance students' attitudes towards creativity, but also students' active learning and exploration (Saxon et al., 2003). Positive effects on students' divergent thinking and creative performance were reported as well (Kashani-Vahid et al., 2017; Kim et al., 2019). A key task for educators is to practice CPS with students from an early age on and to develop a creative habit for life (Kim et al., 2019).

For almost seven years, I worked as a primary school teacher. Fortunately, I was teaching at a school that wanted to stimulate the development of students' CPS by regularly offering CPS activities to students. Nevertheless, I also experienced that teachers struggled to make room for creativity in their lessons. Teachers not only missed the resources to promote creative abilities in students (cf. Davies et al., 2014), they also felt that the increasing weight of standardized testing hampers creativity (cf. Solomon, 2009). When stakes are high, teachers will be more prone to prepare students for what will be tested (Black & Wiliam, 1998). As standardized tests and the vast majority of the lessons focus on knowledge and (sometimes) skills, this is what students will focus their attention and effort on (Van der Vleuten et al., 2000). Teachers' struggle to embed creativity in the curriculum is reflected in the so-called 'creativity gap': the tendency of students to show less creativity inside school than outside school (Runco et al., 2017). Back to the drone: In order to nurture the CPS abilities that the fourth grader in our example has, it is essential to pay attention to CPS from as early on as primary education and to assess it accordingly. Therefore, this thesis focuses on CPS in primary school students and especially on how it can be assessed.

1.1 CPS: A Creative Ability

CPS is regarded as a form of creativity. A widely accepted definition of creativity is:

> *The interaction among aptitude, process and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context (Plucker et al., 2004, p. 90).*

The social context in this definition may refer to a group of eminent artists, but may also be a classroom full of fourth-graders. The product may be a unique piece of art, but it may also be a creative idea to keep ice popsicles frozen. Whereas eminent artists' unique pieces of art are often described as a form of so-called 'Big-C' genius creativity, the creative ideas that solve the ice popsicles problem can be regarded as a form of 'little-c' or everyday creativity (Kaufman & Beghetto, 2009). In creative processes, two main thinking processes can be distinguished: divergent thinking and convergent thinking (Barbot et al., 2016a; De Vink et al., 2022). Divergent thinking includes generating ideas after being confronted with a stimulus (Reiter-Palmon et al., 2019). In traditional divergent thinking tasks, students are, for instance, asked to think of multiple and creative alternative uses for a toothbrush or a paperclip. Convergent thinking, on the other hand, entails the evaluation and selection of ideas with the goal to find the most original and useful ones (Brophy, 2001). Although it is difficult to completely separate these two processes (i.e., one may already apply some form of selection or refinement before expressing an idea), creativity task instructions usually focus on divergent or convergent processes (e.g., De Vink et al., 2022). Original ideas are, by definition, scarce. Because highly original ideas mostly come up during the processes of generating rather than evaluating ideas, previous research mostly focused on divergent thinking. Following creativity theory and the definition of Plucker and colleagues (2004) though, ideas should not solely be original, but also useful (Corazza, 2016). Therefore, both divergent and convergent thinking deserve attention when studying creativity in primary school students.

When teachers do manage to incorporate creativity in their lessons, they often rely on divergent thinking tasks (Sun et al., 2020). However, these tasks (e.g., thinking of alternative uses for a toothbrush) are criticized for having little connection to the problems that students face in everyday life (Piffer, 2012; Zeng et al., 2011). When teachers try to implement creativity in their teaching, they often feel they must abandon the subject matter prescribed in the curriculum which is also often part of the standardized tests (Kaufman et al., 2016a). One way to help teachers to overcome this challenge, is to emphasize a 'both/and approach' instead of this 'either/or approach' (Beghetto, 2013). This both/and approach implies offering activities that cover subject matter and demand creativity instead of offering context-independent creativity tasks next to activities covering curriculum content. In this way, the both/and approach allows teachers to infuse creativity into their existing curriculum. Open-ended exploratory contexts such as CPS tasks focus on both divergent and convergent thinking and fit this both/end approach, because they demand both domain knowledge and general creative processes (Burnard et al., 2006; Poddiakov, 2011). In this way, CPS tasks may contribute to the urgency to embed more flexible ways of dealing with knowledge learned at school (Mayer, 2006) and may counterbalance the argued lack of room for creativity in academic achievement tests (Solomon, 2009).

Multiple scholars have proposed models describing CPS processes that can be applied in practice accordingly (e.g., Altshuller, 1996; Finke et al., 1992; Mumford et al., 1991; Simon, 1969; as cited in Rowe, 1987). Treffinger, Isaksen, and colleagues (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) developed a framework for teaching CPS in education. Their CPS framework includes four interconnected elements that can be applied in diverse problem situations, and that should be regarded as flexible and dynamic in their ordering.

The first element, 'understanding the challenge', focuses on orientation, preparation, and the construction of opportunities to kick-start idea generation while retaining focus (Isaksen et al., 2011). This means that processes of fact finding and problem finding are applied: domain-knowledge is explicitly explored, and the problem at stake is defined. In the second element, 'idea generation', the main goal is to apply divergent thinking and to suggest many, varied and unusual options for responding to a problem. These processes are also referred to as idea finding. The third element, 'preparing for action', includes processes of solution finding: the evaluation of ideas to identify the most creative ones. Because in every CPS process, students tend to switch between elements multiple times while solving a problem (Isaksen et al., 2011; Treffinger et al., 2008), the fourth element, 'planning your approach', is necessary to keep the CPS process on track and therefore includes efforts to plan and monitor.

The CPS model can be regarded as a general tool that can be applied across a variety of problems (Isaksen et al., 2011). Nevertheless, a specific problem situation requires specific knowledge that has to be explored and applied (Reiter-Palmon et al., 2009). Although more research is necessary to determine to what degree CPS is a domain- or task-specific ability, CPS is usually regarded as a set of general routines that are applied in a specific context (Isaksen et al., 2011). Examples of domains that CPS could be applied in are the scientific, interpersonal, and entrepreneurial domains (Kaufman, 2012; Oral et al., 2007).

Next to the four interconnected elements, four indicators are used to determine the creativity of the ideas that have been generated: fluency, originality, completeness, and practicality. The number of generated ideas to solve the problem is often taken as an indicator of CPS fluency. In other words, this indicator describes how fluent a student is in generating ideas (Isaksen et al., 2011; Reiter-Palmon et al., 2019). The second indicator, originality, describes how novel or unique ideas are. In line with the definition of Plucker and colleagues (2004), creativity does not solely require the production of unique ideas; also, some form of ‘appropriateness’ is required. For CPS in particular, this means that a proposed solution should also be detailed, solve the problem at stake, and ‘work’ in practice (Corazza, 2016; Reiter-Palmon et al., 2009). In other words, the solution should be complete and practical.

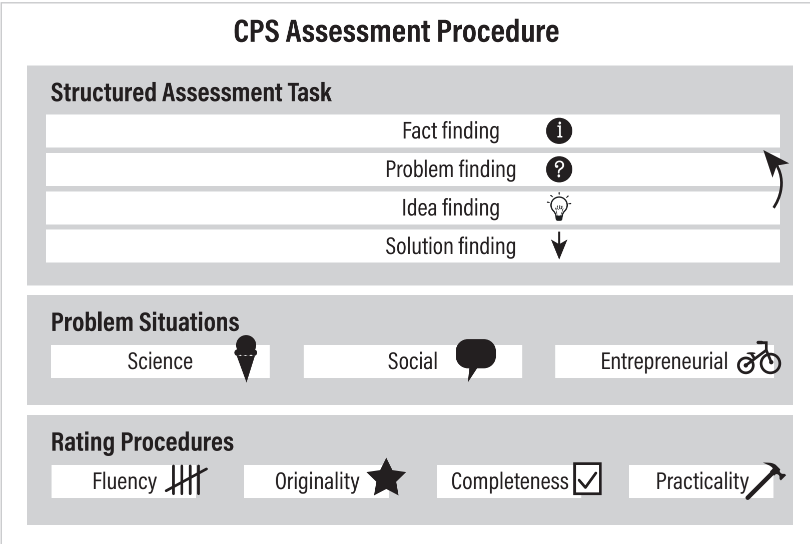
1.2 The Assessment of CPS

Understanding what students’ creative abilities are is essential for nurturing these abilities in the classroom (Benedek et al., 2016; Bolden et al., 2020). This is why CPS should not only be promoted, but also assessed. Although the four CPS indicators are well defined, assessing CPS is still a challenge. Tasks that demand creativity, including CPS tasks, tend to produce different outcomes due to a variety of factors including differences in prior knowledge, task characteristics, and motivational factors (Amabile & Pillemer, 2012; Conti et al., 1996; Piffer, 2012; Reiter-Palmon et al., 2009). A very social student might, for instance, be more able to show his or her creativity on an interpersonal task compared to a science-oriented student, whereas a science-oriented student might be better able to solve a science problem. So not only general CPS ability but also differences in subject knowledge, previous experiences with the problem, and personal preferences tend to produce differences in the scores on CPS tasks. This illustrates the potential domain- or task-specific nature of CPS. To deal with this variability across creativity tasks, complete creativity assessment procedures are developed that include multiple tasks and detailed rating procedures (e.g., EPoC; Barbot et al., 2016a). In this way, a more reliable picture of a student's creative abilities is obtained. Such an assessment procedure to assess CPS in primary school students does not yet exist. As a consequence, little is known about the nature of CPS in children and how it may be assessed.

In order to nurture CPS in the classroom, teachers must be able to assess CPS as well. Unfortunately, multiple authors found that the accuracy of teacher ratings of students' creativity is generally low and that teachers tend to be biased by gender and school functioning in their judgements (Beghetto et al., 2011; Gralewski & Karwowski, 2013; Urhahne, 2011). Furthermore, the variability in creativity scores across tasks mentioned earlier, hampers a standardized assessment of CPS (Harris, 2016; Reiter-Palmon et al., 2009). More insight is needed in whether and in what way teacher assessments could be used to gain a valid estimate of students' CPS abilities. Within the studies described in this thesis, I therefore also studied teachers' assessments of CPS and absolute and relative ways of scoring, to explore how the assessment of CPS could be transferred to educational practice.

Figure 1

Graphical representation of the CPS assessment procedure



1.3 This Thesis

For this thesis, a CPS assessment procedure was developed and tested. This assessment procedure includes a structured CPS task that focuses on fact-finding, problem-finding, idea-finding, and solution-finding. Three problem situations stemming from the science, social and entrepreneurial domain were included. Rating procedures focusing on the four indicators to assess the creativity of the ideas that students generate were developed as well (Figure 1).The ultimate goal of the studies described in this thesis is to contribute to the assessment of CPS in the primary school context and, in this way, enable teachers to nurture CPS in students by assessing CPS regularly. In order to do so, CPS and this assessment procedure should first be studied more thoroughly in the primary school context. This project therefore had two main goals: (1) to gain insight into the nature of CPS in primary school students, and (2) to formulate implications for the assessment of CPS.

Below, the separate chapters of the current thesis are introduced.

Chapter 2

When CPS is applied in education, generating creative ideas is often taught in isolation, rather than also including processes such as exploring knowledge, defining the problem, and comparing ideas to identify the most creative ones. In the study described in this first chapter, me and my colleagues therefore evaluated whether the different steps of fact finding, problem finding, idea finding and solution finding in a structured assessment task 'work'. That is, we explored to what degree these steps contribute to eventual CPS outcomes. For this purpose, a task was administered two times in a sample of six 4th- and 5th-grade classes (N = 137 students). We examined whether the two processes of the 'understanding the challenge' element – fact-finding and problem-finding – contribute to the eventual creativity of the generated ideas. Additionally, we studied whether solution finding is doable for elementary school students, by applying the top-scoring method (Benedek et al., 2013): students themselves selected their three most creative ideas and it is examined how students selected these ideas. The two research questions were :

Question 1.1: Are successful fact-finding and problem-finding positively associated with idea finding of primary school students across CPS tasks?

Question 1.2: Can primary school students select ideas scoring high on multiple aspects of creativity across CPS tasks, or do they ignore some aspects, such as originality?

Chapter 3

In chapter 3, a method is described to study creativity assessment procedures more comprehensively. The theory that underpins this method, Generalizability Theory, (i.e., a statistical theory and method to study the generalizability of assessments; Brennan, 2010) was applied to the same CPS data as in Chapter 2 to illustrate how this method could be applied in creativity research. To shed light on the representativeness of CPS tasks, we reviewed whether CPS performance, as measured with a CPS task, can be generalized across problem situations. We studied the generalizability for both absolute and relative scores. Information is also gathered on how raters and tasks impact CPS scores. The research question was:

Question 2.1: Can CPS performance, as measured with the present approach, be generalized across problem situations?

Chapter 4

In Chapter 4, we sought further insight into the nature of CPS processes in primary school students. Both the CPS process (as described by the model of Isaksen and colleagues, 2011) and the product (the set of creative ideas to solve a problem, scored on the four indicators; see also Figure 1) were studied. For this purpose, we conducted two separate studies. First, we qualitatively studied how students solve problems creatively with the task with a small-scale think-aloud study (N = 13). In a second study, the task was administered three times in a large sample of 25 classes (N = 594), and two raters rated the CPS performance of the students. We evaluated whether the CPS indicators can be aggregated across tasks. The relations between these aggregated indicators are compared to relations found in earlier studies to determine whether CPS in primary school students resembles CPS in adult samples. In addition, the students' CPS performance is related to outcomes of a divergent thinking task and to academic achievement. The three research questions were:

Question 3.1: To what extent do the CPS elements appear when primary school students solve problems creatively?

Question 3.2: How are the CPS indicators interrelated in primary school students?

Question 3.3: How do these CPS indicators relate to outcomes from a divergent thinking task and academic achievement?

Chapter 5

Especially when the aim is to assess CPS in the classroom, teacher ratings of CPS should give a reliable estimation of students' CPS skills. Therefore, in Chapter 5, we studied how teachers assess the CPS performance of students. For this purpose, teachers (N = 26) assessed the CPS performance of their students (N = 610; same sample as Chapter 4 with one extra class). These teacher assessments were compared with the assessments of trained raters. Additionally, potential biases in teachers' assessments of CPS were explored by studying whether the gender and academic achievement scores of the students informed the assessments of the teachers. The research question was:

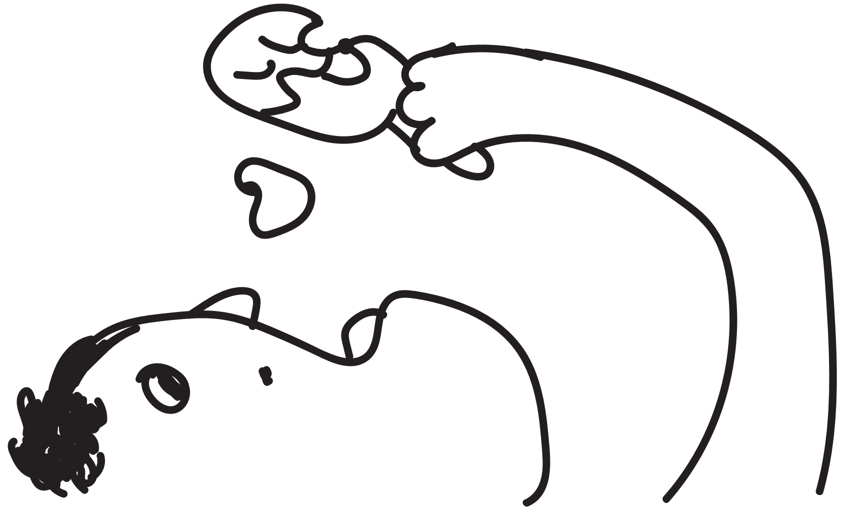
> *Question 4.1: Can elementary school teachers assess students' creative problem solving abilities?*

Chapter 6

This final chapter presents a summary and discussion of the main findings of the collection of studies that I conducted with my colleagues. Here, I also draw more general conclusions about the nature of CPS in primary school students and what the results of the studies imply for the assessment of CPS. Directions for future research and recommendations for teachers will also be shared in this final chapter.

2

CREATIVE PROBLEM SOLVING IN PRIMARY EDUCATION: EXPLORING THE ROLE OF FACT FINDING, PROBLEM FINDING, AND SOLUTION FINDING ACROSS TASKS



"Eat the ice pops upside down, so the drops fall into your mouth."

This chapter is based on:
Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & van Tartwijk, J. (2020). Creative problem solving in primary education: Exploring the role of fact finding, problem finding, and solution finding across tasks. Thinking Skills and Creativity, 37, 100665. https://doi.org/10.1016/j.tsc.2020.100665

Acknowledgement of author contributions:
JvT acquired funding for the study; MvH and EK designed the study; MvH collected the data; MvH and EK planned the data analyses; MvH analysed the data; MvH drafted the manuscript; all authors contributed to critical revision of the manuscript; EK, TM and JvT supervised the study.

Supplementary material can be found at: https://doi.org/10.34894/RGVQXK

Abstract

Interest in fostering creative problem solving (CPS) from primary education onwards is growing. However, embedding CPS in Education seems to be a challenge. One problem is that generating creative ideas (idea finding) is often taught in isolation, rather than also including processes such as exploring knowledge (fact finding), defining the problem (problem finding) and comparing ideas to identify the most creative ones (solution finding). In the current study, we prepared CPS tasks for primary education that represent this more complete CPS model and studied whether successful fact finding and problem finding were positively associated with the creativity of the ideas found. Additionally, we studied whether solution finding is doable for these young students and how they select the most creative ideas. Bayesian analyses indicated a positive association of fact finding and problem finding with the number of ideas generated and the originality of these ideas. Furthermore, problem finding seemed to be positively associated with the completeness of ideas, whereas fact finding seemed not. We also found that primary school students were able to identify their most creative ideas. Students did not seem to undervalue certain aspects of creativity when applying solution finding. Our results indicate that when aiming for more and original solutions, teachers could embed fact finding and problem finding in their CPS teaching practices. Our results also indicate primary school students are able to recognize creativity.

2.1 Introduction

Modern society requires people to be able to solve problems in a creative way (Craft, 2011). As such, educational systems need to produce creative problem solvers that try to understand everyday challenges, generate multiple creative ideas and select the most creative ideas to put into practice (Isaksen et al., 2011). An idea is seen as creative when it is original, well-thought out (i.e., complete) and transferrable to practice (Corazza, 2016; Okuda et al., 1991; Reiter-Palmon et al., 2009). Multiple scholars created frameworks to describe the CPS process and to support finding these creative ideas (e.g., Altshuller, 1996; Finke et al., 1992; Mumford et al., 1991; Simon, 1969; as cited in Rowe, 1987). Among others, Treffinger and colleagues (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) continued to develop this framework. Their CPS model includes three main components: (1) understanding the challenge (2) generating ideas, and (3) preparing for action. The goals of the understanding the challenge stage are orientation, preparation, and the construction of opportunities to kick-start idea generation and to retain focus at the same time (Isaksen et al., 2011). Within this first stage, students apply fact finding by exploring and defining their knowledge on the problem. In addition, they apply problem finding by identifying the problem at stake. In the second idea generation stage, students are asked to diverge and come up with creative ideas to solve the problem (Isaksen et al., 2011). This stage is also called idea finding. In the last preparing for action stage, students need to evaluate their ideas and identify their most creative ones (Isaksen & Treffinger, 2004). This is also called solution finding and is seen as a last step before the ideas can be transferred to practice (Isaksen et al., 2011).

Interest in CPS from primary education onwards is growing, because it can be easily connected to the problems a student may face in daily life as well as to factual knowledge learned in school (Runco & Acar, 2012). Besides this, engaging in CPS showed to be beneficial for primary school students' divergent thinking, attitudes towards creativity, active learning and exploration (Kashani-Vahid et al., 2017; Saxon et al., 2003). However, embedding CPS in education seems to be a challenge. Only in a few studies, the translation of the full CPS process to educational practice and to primary education in particular was investigated (e.g., Arreola & Reiter-Palmon, 2016; Okuda et al., 1991). When teaching CPS, teachers often focus on divergent thinking and idea finding (Cropley, 2006; Piffer, 2012). This means students are often immediately asked to think of multiple creative ideas to solve a problem at the expense of processes demanding both convergent and divergent thinking such as fact finding, problem finding, and solution finding processes (Isaksen et al., 2011). We do however not know whether fact finding and problem finding may help primary school students to think of more creative ideas in the idea finding stage. Additionally, we do not know whether solution finding is doable for these young students and how they select the most creative ideas. CPS outcomes showed to differ across CPS tasks suggesting that CPS processes and how they are interrelated might differ across CPS tasks as well (Reiter-Palmon et al., 2009). In the current study, we therefore prepared two CPS tasks for primary school students that next to idea finding explicitly embedded and scaffolded fact finding, problem finding, and solution finding processes. We explored whether successful fact finding and problem finding are, as the CPS model would predict, positively associated with idea finding and whether primary school students are able to apply solution finding by identifying their most creative ideas.

Fact finding

In the fact finding stage, students are asked to explore their knowledge on the problem and to describe or list all the knowledge elements they can think of (Isaksen et al., 2011). Fact finding is of particular importance because factual knowledge can be regarded as a precondition for inferencing and high-level reasoning (Yekovich et al., 1991). Besides this, it is seen as the source of creative ideas (Cropley, 2006). Fact finding may especially affect the ability to come up with many ideas (i.e., fluency), because listing the knowledge connected to a problem may highlight more solution opportunities (McCaffrey, 2016). Barak (2013) found that a systematic search of the knowledge involved was beneficial for high schoolers' creative outcomes. For undergraduates, length of time spent on factual information was also positively associated with generating ideas of higher quality and originality (Mumford et al., 1996a). Length of time spent on factual information also mediated the relationship between problem construction engagement and creativity, next to the quantity and the breadth of information viewed (Harms et al., 2018). Although there is no evidence available that fact finding also helps primary school students in generating ideas, in line with findings from Barak (2013) and Mumford and colleagues (1996a) we hypothesize that successful fact finding is positively associated with idea finding. We especially expect fact finding to be positively associated with the number of ideas a student comes up with (i.e., fluency scores).

Problem finding

Students are asked to think about the problem at hand and describe it in the problem finding stage (Isaksen et al., 2011). This could help in providing focus when applying idea finding. Mumford and colleagues (1996b) found that problem finding ability explained differences in CPS performance of high schoolers, even after controlling for general intelligence and divergent thinking ability. In primary education, problem finding was beneficial for students when solving novel problems in mathematics (English, 1997) and was predictive of creative accomplishments (Ma, 2009; Okuda et al., 1991). However, problem finding was also a challenge for primary school students (Van Harpen & Sriraman, 2013). Nonetheless, the positive effects of problem finding for this group are not completely clear. Arreola and Reiter-Palmon (2016) found a positive association of problem finding and idea finding outcomes of undergraduates, but this association was different across different problem tasks. For one task, problem finding quality predicted idea finding quality, whereas for another task the quality of the problems found predicted both idea finding quality and originality. Although several researchers asked participants to come up with multiple possible problems (e.g., Arreola & Reiter-Palmon, 2016; Mumford et al., 1996b), we asked primary school students to identify one single problem to keep our CPS task as simple as possible. We hypothesize problem finding is positively associated with idea finding, but especially for measures of completeness because we think the provided focus helps students to develop well-rounded solutions. We will explore whether this hypothesis is true across two CPS tasks to screen for potential task differences.

Solution finding

During solution finding, students are asked to evaluate their own creative ideas and select the most creative ones. In the past years, the number of studies investigating the identification of creative ideas as a feature of creative metacognition (CMC) in education has grown (e.g., Benedek et al., 2016; Grohman et al., 2006; Kaufman et al., 2016b; Pretz & McCollum, 2014; Silvia, 2008; Sternberg, 2012). In primary education, Runco (1991) found that students' evaluative abilities of divergent thinking tasks can be measured and that evaluative skill is related to divergent thinking ability. However, the popularity of ideas was more accurately judged than the creativity of ideas. Grohman and colleagues (2006) conducted an exploratory study on whether primary school students' self-ratings of creativity aligned with external ratings of creativity on a visual, verbal, and scientific task. Results showed students were able to differentiate their performance across creativity domains and across quality levels. Besides this, their scores were positively associated with creativity scores given by experts. Within this study of Grohman and colleagues (2006) creativity of the ideas was reviewed as a whole. The aspects of creativity can however be separated to explore whether specific creativity aspects are valued more in the solution finding process than others. This will give us more insight in whether all aspects are valued by students when solution finding is applied in education and will gives us more detailed advice on how to train solution finding. Rietzschel and colleagues (2010) found in their study with undergraduates that when doing CPS, even with explicit training, it seemed to be difficult to both appreciate originality and effectiveness of their ideas. Blair and Mumford (2007) found in their study with undergraduates that evaluation criteria and time pressure influenced what creativity aspects were valued most by students. The question is whether primary school students are able to select ideas scoring high on all three aspects of creativity across CPS tasks, or ignore certain aspects such as originality. We hypothesize students are able to apply solution finding by selecting the most creative ideas. Because less is known about what aspects of creativity are valued most by primary school students, we will not formulate specific follow-up hypotheses here but explore whether the separate aspects are valued or undervalued.

The present study

To enhance CPS in everyday educational practice, we need to gain more insight in whether the different CPS stages help these young students to generate and recognize creative ideas. In the current study, we therefore prepared CPS tasks for primary education that represent the CPS model (Isaksen et al., 2011) and embed fact finding, problem finding, and solution finding alongside idea finding. Within this explorative study, we will answer two sets of questions:

> *Question 1.1: Are successful fact finding and problem finding positively associated with idea finding of primary school students across CPS tasks?*
> *Question 1.2: Can primary school students select ideas scoring high on multiple aspects of creativity across CPS tasks, or they do they ignore certain aspects such as originality?*

Answering these questions will not only give us insight in how we can foster CPS from primary education onwards, it will also help us to develop interventions to enhance primary school students' ability to generate and recognize creative ideas across problem situations.

2.2 Method

Participants

Six classes of 4th and 5th grade students (N = 148; mean age = 10.38, SD = 0.72; 72 girls, 76 boys) of three Dutch primary schools (one urban, one sub-urban, one village school) participated voluntarily in the study. Parents were informed about the study and all messages to parents included an anonymity statement. Parents and students were informed and told that students would be asked for consent and could withdraw from the study at any point. Five students did not participate due to illness on the day when the students completed the tasks. Before the start of the CPS tasks, three students asked whether they could withdraw, and were as such excluded from the study. All remaining students (N = 140) were randomly divided in groups of eight to eleven students and were individually seated outside the class in a quiet study area. Tasks of three students turned out to be unscorable in the rating phase. As such, data from these students was excluded as well (N = 137).

Procedure

Every student completed two CPS tasks individually. Because the science and interpersonal domains were found as different creativity domains in several studies (Kaufman, 2012; Kaufman et al., 2009; Oral et al., 2007), we selected one problem situation for each of these domains from Treffinger's practice problems for many ages (2000). With vignette theory (Poulou, 2001), we modified the problem situations to fit our research purposes and the age group. Vignettes are short descriptions of hypothetical situations (or persons) that include the required information for the participants to reason with (Poulou, 2001). As with vignettes, the problem situations described hypothetical scenarios which, although they were very realistic, did not involve the respondent personally. They were constructed to attract the interest of the students and stimulate their imagination and written in third person to avoid personal biases such as personal experiences with the problem. The problem situations were open and non-directive since they enable the participant to form his/her own interpretation of the described situation, important for creativity. Furthermore, they were clearly written to make sure they were easily understood by the respondents. The problem situations were discussed with two teachers to check whether they were fit for purpose. Two sentences were shortened and one sentence was rewritten to improve readability. The final problem situations included a short story about the problem, and were presented written on paper as well as read to the students by the researcher. The science problem described Lisa and Tom buying ice cream on a hot day (Figure 1). The social problem referred to a classroom situation, in which Simon gets distracted by his friend Julian all the time (Figure 2).

Figure 1

Problem situation 1, stemming from the science domain.

The weather is very hot. Lisa and Tom feel like eating ice cream. They cycle to the supermarket and buy a colored cardboard box with ice pops. However, when they get home and open the wrapper, the ice pops appear to be completely soft. The ice drips slowly down the stick. The ice pops started to melt in the box!

Figure 2

Problem situation 2, stemming from the social domain.

Simon sits next to his friend Julian in class. Julian likes to talk to Simon and often bothers him when he is busy with his work. Sometimes Julian distracts Simon, making him miss an important part of the lesson. Today, Simon does not get his work done once again because Julian interferes with him. However, Julian is also a very good friend of Simon, and he really likes hanging out with him.

Students received a short explanation of the CPS steps involved during the first CPS task (science). First, students applied fact finding by listing the knowledge elements they could think of (e.g., the box, the ice, the distance, the temperature; cf. Barak, 2013). Students were told they could list as many elements as they liked. Then, students applied problem finding by writing down the problem statement in the form of a question. Subsequently, students received ten minutes to list as many different and original ideas to solve the problem as they could. Because instructions influence creative outcomes (Nusbaum et al., 2014; Runco et al., 2005), the students were explicitly asked to come up with ideas nobody else would think of. Besides this, students were not notified about the time. Although the authors state that the CPS process can have different starting points and routes (Isaksen et al., 2011), we choose to apply these components as stages in a particular order to be able to scaffold the CPS process of the young creative problem solvers. However, to stimulate the use of knowledge and the problem when applying idea finding, students were stimulated to look back at their first stages to see what they could use in the construction of more ideas.

In the final step of the procedure, students chose their three most creative ideas. Silvia and colleagues (2008) proposed a method that facilitates solution finding: top-scoring. Within this procedure, participants are asked to choose a specific number of most creative ideas, which are then considered for scoring or evaluation (e.g., Benedek et al., 2013). To come to this top three within our study, students were asked to take originality, completeness and practicality into account: The administrator stated: "A creative idea is an original idea nobody has thought of before, it is complete and solves the problem and can be easily transferred to practice." The students were asked to put one number one, one number two, and one number three at their selected ideas to highlight their top 3.

Measures

Two raters (trained teachers and involved in educational research; one graduate, one post-graduate) were used to rate the CPS tasks. They received at least 8 hours of training to understand the CPS tasks, the CPS concepts and the ratings schemes involved. Pilot data of 20 students was rated and discussed to establish sufficient inter-rater agreement. Because every student received ratings from these two raters and scores were averaged, the two-way mixed average score Intra Class Correlation (ICC) was calculated to check the inter-rater agreement.

Fact finding

Fact finding was scored as the number of different knowledge elements listed. Elements that were mentioned twice and elements that could not be considered as knowledge elements (e.g., "this is a problem"), were excluded from scoring.

Problem finding

The quality of the identified problem was evaluated using a modified version of the consensual assessment technique (CAT; Amabile, 1996). Problem finding was defined as the quality of the identified problem and scored by the two raters (ICC = .84 indicating excellent agreement; Cicchetti, 1994) on a five-point scale. For instance, when the problem identified was not or marginally related to the situation involved, it received 0 points. (e.g., warmth). When the problem identified was stated in the form of a question, was complete, accessible for solution and shaped the focus of the CPS process, it received 4 points (e.g., How can we keep ice from melting on a hot day, when we are buying it in the supermarket?).

Idea finding

Four idea finding measures were used: fluency, originality, completeness, and practicality.
To assess fluency, the method commonly used to score divergent thinking tasks was applied (Runco & Acar, 2012). The fluency score consisted of the total number of different ideas listed. Ideas that could not be interpreted or were listed twice were excluded.
Originality was rated with a rating scheme by the two raters (ICC = .83, indicating excellent

agreement; Cicchetti, 1994) on a five-point Likert scale. If the idea was very predictable and commonly known, it received 0 points (e.g. use a cool bag). If the idea clearly reflected an imaginative approach and was completely new, it received 4 points (e.g., make an umbrella that protects you from the sun and has little fans built-in that blow cold air).
Completeness was rated by the two raters (ICC = .67, indicating good agreement; Cicchetti, 1994) on a five-point Likert scale. If the problem at stake was ignored or just repeated in the idea, it received 0 points (e.g., don't eat ice-cream, make sure it does not melt). If multiple steps towards a solution were included in the idea, it received 4 points (e.g, put a fridge on your bike that is powered by a dynamo, put the ice in when you leave the supermarket and cycle home).
Practicality was rated by the two raters (ICC = .90, indicating excellent agreement; Cicchetti, 1994) on a 5-point Likert scale. If the idea was impossible in practice, it received 0 points (e.g., go to a wizard to refreeze the ice-cream with magic). If the idea could be transferred to practice right away, it received 4 points (e.g., buy new ice-cream and eat the ice-cream just outside the store on a bench.) The scores for originality, completeness, and practicality of the entire set of generated ideas were averaged across raters to get scores for idea finding next to the fluency scores.

Solution finding

By averaging the scores for originality, completeness and practicality across the top-scores, we gained measures for solution finding. Because the number of ideas was too low (<4) for some students, we compared the top-1 and the top-3 score with the mean score across all ideas within these samples. Findings of the top-1 and top-3 analysis were compared (Table 3).

Analysis

Within this study, Bayesian statistics were used instead of conventional Null-hypotheses tests (Van de Schoot et al., 2014). Bayesian statistics were used because we held specific, informed hypotheses (e.g., that problem finding has an even larger influence on completeness than fact finding) rather than uninformed, null hypotheses. Furthermore, Bayesian statistics are less sensitive for problems with normality (e.g., relatively scarce original ideas) and eliminate the risk of alpha inflation when conducting multiple tests.

Bayesian statistics give researchers the chance to include prior knowledge when conducting their analyses (Klugkist et al., 2011). Within a Bayesian analysis, informed hypotheses are compared with each other and with an unconstrained hypothesis to correct for poorly chosen hypotheses (Kluytmans et al., 2012). The analyses produce Bayes Factors (BFs), indicating the support for the hypothesis from the data. A BF larger than 1 indicates marginal support for the hypothesis, BFs from around 3 can be regarded as positive evidence, and when BFs of 6 and bigger are found, evidence can be regarded as strong (Hoijtink et al., 2016; Kass & Raftery, 1995). BFs can be divided by one another, to calculate the likeliness of one hypothesized model compared to another. In addition, posterior model probabilities (PMPs) can be calculated by dividing the BF of the chosen model by the sum of BFs of all models of interest. This PMP represents the relative support for a specific hypothesis within a set of competing hypotheses (Klugkist et al., 2011).

The analyses were performed with the program BIEMS (Mulder et al., 2012). Because no previous outcomes are available, an informed prior could not be set. That is why we used the standard prior in BIEMS. This approach is based on using a subset of the data for automatic prior specification (Mulder & Wagenmakers, 2016; Nathoo & Masson, 2016).

2

Question 1.1: Fact finding & problem finding.
To investigate whether fact finding and problem finding are positively associated with idea finding, we applied Bayesian Linear Regression. We regressed our four measures for idea finding (fluency, originality, completeness and practicality) on fact finding and problem finding. Because we hypothesized that fact finding and problem finding were beneficial for idea finding, our hypothesized H1 model included a positive influence of both predictors (fact finding = ff, problem finding = pf): H1: $\beta_{ff} > 0, \beta_{pf} > 0$. For fluency, we expected fact finding to have an even larger influence than problem finding, so a second hypothesized model was added: H2a: $\beta_{ff} > \beta_{pf} > 0$. For completeness, we expected problem finding to be most beneficial, so here we also added a second hypothesized model H2b: $\beta_{pf} > \beta_{ff} > 0$. These hypotheses were compared with the unconstrained model (H_{unc}: β_{ff}, β_{pf}), which assumes no specific direction. The BFs will be interpreted alongside the standardized regression coefficients (betas) to draw conclusions on whether fact finding and problem finding are positively associated with idea finding. Findings were compared across the two tasks to screen for potential task differences.

Question 1.2: Solution finding
To investigate whether students were able to identify creative ideas, we conducted a Bayesian paired sample t-test for the top-1 and the top-3 score for originality, completeness and practicality of the ideas together (multivariate). Because we were interested in whether students are in general able to identify creative ideas, we hypothesized all top-scores for the separate aspects (originality, completeness, practicality) were larger than the overall scores, resulting in the hypothesized model H1: $\mu_{top\text{-}ori} > \mu_{all\text{-}ori}$, $\mu t_{op\text{-}com} > \mu_{all\text{-}com}$, $\mu_{top\text{-}pra} > \mu_{all\text{-}pra}$.

Because we were interested to see on which CPS aspects students tend to base their selection of top ideas, we compared two additional competing hypothesized models per aspect: (1) H1: Students choose ideas with a relatively high score on originality/completeness/practicality and as such are able to detect this aspect ($\mu_{top} > \mu_{all}$), and (2) H2: Students undervalued a certain aspect when they choose their best ideas ($\mu_{top} < \mu_{all}$). These models were compared to the unconstrained model (H_{unc}) assuming no specific direction: students tended to choose ideas neither high nor low on creativity (i.e., appeared to pick an idea randomly; μ_{top}, μ_{all}).

2.3 Results

Question 1.1: Fact finding & problem finding
To investigate whether fact finding and problem finding are positively associated with idea finding, we regressed our four measures for idea finding (fluency, originality, completeness and practicality) on fact finding and problem finding. Descriptive statistics of the dependent and independent variables are provided in table 1.

Table 1

Descriptive Statistics of Fact Finding, Problem Finding and Idea Finding Measures (N = 137).

	Science CPS task				Social CPS task			
	M	**SD**	**Min**	**Max**	**M**	**SD**	**Min**	**Max**
Fact finding	5.77	2.61	0	16	5.79	3.26	0	14
Problem finding	2.40	0.84	0	4	1.92	0.97	0	4
Idea finding								
Fluency	8.29	3.79	2	21	6.67	3.70	2	18
Originality	0.56	0.43	0.00	1.90	0.37	0.42	0.00	1.67
Completeness	2.42	0.46	1.30	3.90	2.31	0.34	1.63	3.30
Practicality	2.70	0.57	0.80	3.63	3.25	0.56	1.22	4.00

For fluency, the first hypothesized model, including a positive effect of fact finding and problem finding ($\beta_{ff} > 0$, $\beta_{pf} > 0$), resulted in BFs of 3.09 (science task) and 2.91 (social task; Table 2), which means this model was about three times more likely to be true than the unconstrained model. However, the additional H2a model, which included a positive effect of both as well as a larger effect of fact finding ($\beta_{ff} > \beta_{pf} > 0$), resulted in BFs of 6.12 (science task) and 5.21 (social task; Table 2). This indicated that this model was about 5 to 6 times more likely to apply than the unconstrained model and about 2 times more likely than the H1 model. The positive posterior betas of fact finding (.37 and .24) and problem finding (.06 and .05; Table 2) illustrate the relevance of this model. Overall, these results indicated a positive association of both fact finding and problem finding with fluency, with the largest positive association of fact finding. A similar trend occurred across the two CPS tasks.

For originality, the proposed hypothesized model, which included a positive effect of fact finding and problem finding ($\beta_{ff} > 0$, $\beta_{pf} > 0$), resulted in a BF of 3.79 (science task) and 3.70 (social task; Table 2). This means that both models are almost four times more likely than the unconstrained model respectively, indicating a positive association of fact finding and problem finding with originality. Again, this is illustrated by the positive posterior betas. A similar trend occurred across the two CPS tasks.

For completeness, the H1 model, which included a positive effect of fact finding and problem finding ($\beta_{ff} > 0$, $\beta_{pf} > 0$), resulted in BFs of 0.84 (science task) and 0.79 (social task; Table 2), implying that this model was not supported by the data. However, the additional H2b model including a positive effect of both, plus a larger effect of problem finding ($\beta_{pf} > \beta_{ff} > 0$) resulted in BFs of 2.61 (science task) and 2.55 (social task; Table 2) implying marginal support from the data. This indicated that, based on the data, this model was about two-and-a-half times more likely than the unconstrained model and about three times more likely than the H1 model. This seems to indicate that problem finding tends to have a larger positive association with completeness than fact finding. This is illustrated by the posterior betas of -.07 for fact finding and 0.15/0.18 for

problem finding (Table 2). Findings were again similar across the two tasks.
For practicality, the hypothesized model including a positive effect of fact finding and problem finding ($\beta_{ff} > 0$, $\beta_{pf} > 0$) resulted in in a BF of 0.11 (science task) and 0.62 (social task; Table 2). This means this model did not receive support from the data, especially for the science problem. This is illustrated by the negative betas of fact finding and problem finding for the science task, and the negative beta of fact finding for the social task.

Table 2
Results Bayesian Regression Analyses (N = 137).

		Science CPS task				Social CPS task			
	Hypotheses	β_{ff}	β_{pf}	**BF**	**PMP**	β_{ff}	β_{pf}	**BF**	**PMP**
Fluency	H_{unc} .37	.06	1.00	.10	.24	.05	1.00	.11	
	H1: $\beta_{ff} > 0$, $\beta_{pf} > 0$			3.09	.30			2.91	.32
	H2: $\beta_{ff} > \beta_{pf} > 0$			6.12	.60			5.21	.57
Originality	H_{unc} .14	.10	1.00	.23	.14	.13	1.00	.22	
	H1: $\beta_{ff} > 0$, $\beta_{pf} > 0$			3.33	.77			3.56	.78
Completeness	H_{unc} -.07	.18	1.00	.22	-.07	.15	1.00	.23	
	H1: $\beta_{ff} > 0$, $\beta_{pf} > 0$			0.84	.19			0.76	.18
	H2: $\beta_{pf} > \beta_{ff} > 0$			2.61	.59			2.55	.59
Practicality	H_{unc}	-.14	-.09	1.00	.81	-.07	.08	1.00	.62
	H1: $\beta_{ff} > 0$, $\beta_{pf} > 0$			0.11	.09			0.62	.38

Note: β_{ff} = beta for fact finding, β_{pf} = beta for problem finding, BF = Bayes Factor, PMP = Posterior Model Probability. H_{unc} hypothesized no specific direction, H1 ($\beta_{ff} > 0$, $\beta_{pf} > 0$) hypothesized a positive influence of both fact finding and problem finding, H2 for fluency ($\beta_{ff} > \beta_{pf} > 0$) hypothesized a positive influence of both fact finding and problem finding, with the highest influence of fact finding. H2 for completeness ($\beta_{pf} > \beta_{ff} > 0$) hypothesized a positive influence of both fact finding and problem finding, with the highest influence of problem finding.

Question 1.2: Solution finding

Data of thirteen students were excluded due to a missing or unreadable top rating. These students varied in terms of gender, age and class. This resulted in a top-1 sample of 124 students. Because for the top-3 analysis a fluency score of at least 4 ideas for both tasks was necessary because otherwise the students could not make a selection, 25 students needed to be excluded from this analysis. Therefore, for the top-3 analysis, the sample consisted of 99 students. Posterior sample means and variances for the two different samples are provided in Table 3.

Table 3
Posterior Sample Means and Variances of Top-1 and Top-3 Scores and Overall Scores on Originality, Completeness, and Practicality.

	Science CPS Task				**Social CPS Task**			
	M (SD) Top-1	**M (SD) All**	**M (SD) Top-3**	**M (SD) All**	**M (SD) Top-1**	**M (SD) All**	**M (SD) Top-3**	**M (SD) Al**
Originality	0.66 (0.01)	0.57 (0.00)	0.62 (0.00)	0.57 (0.00)	0.47 (0.00)	0.37 (0.00)	0.47 (0.00)	0.40 (0.00)
Completeness	2.63 (0.00)	2.43 (0.00)	2.54 (0.00)	2.42 (0.00)	2.41 (0.00)	2.31 (0.00)	2.35 (0.00)	2.29 (0.00)
Practicality	2.81 (0.00)	2.69 (0.00)	2.77 (0.01)	2.66 (0.00)	3.27 (0.00)	3.24 (0.00)	3.21 (0.01)	3.20 (0.00)

Note: Top-1 N = 124; Top-3 N = 99.

To see whether the students were able to apply solution finding and identify the most creative ideas, we conducted a Bayesian paired sample t-test for the top-1 and the top-3 score for the aspects together (overall creativity; Table 4). The hypothesized model (H1: $\mu_{top\text{-}ori} > \mu_{all\text{-}ori}$, $\mu_{top\text{-}com} > \mu_{all\text{-}com}$, $\mu_{top\text{-}pra} > \mu_{all\text{-}pra}$) assuming students were able to identify their most creative ideas, indicated by relatively high top-scores compared to the overall scores on all three aspects, resulted in BFs ranging from 4.40 (social task, top-3) to 7.06 (social task, top-1; Table 4). This indicated our hypothesis is supported.

BFs coming from our analyses on the separate aspects are also provided in Table 4. The BFs larger than 1 (but lower than 3) for all three aspects indicated marginal support for our hypothesis that students valued the aspect (H1: µtop > µall). The BFs smaller than 1 for all three aspects indicate there is no support from the data for our model proposing student undervalue specific aspects (H2: $\mu_{top} < \mu_{all}$). For originality, the H1 model is almost ten times more likely than the H2 model. Here, we see a similar trend across the two tasks. For completeness, the H1 model (BFs around 2) is from 32 to at least 200 times more likely than the H2 model (BFs ranging from 0.00 to 0.06). Again, there was a similar trend across the tasks, although the support for H1 in the science task is even more convincing. For practicality, the H1 model received about 13 times more support than H2 for this task. For the social task, there was only about 1.5 times more support for H1 compared to H2. In other words, for all three aspects there is more support for the H1 hypothesis. Compared to the H2 hypothesis, the support for H1 is largest for completeness.

2

Table 4
Results Bayesian (Multivariate) T-tests.

		Science CPS Task				Social CPS Task			
		Top-1 (N=124)		Top-3 (N=99)		Top-1 (N=124)		Top-3 (N=99)	
	Hypotheses	BF	PMP	BF	PMP	BF	PMP	BF	PMP
Creativity	H_{unc}	1.00	.13	1.00	.18	1.00	.12	1.00	.19
	H1: μ_{top}-ori > μ_{all}-ori, μ_{top}-com > μ_{all}-com, μ_{top}-pra > μ_{all}-pra	6.64	.87	4.71	.82	7.06	.88	4.40	.81
Originality	Hunc	1.00	.33	1.00	.33	1.00	.33	1.00	.33
	H1: $\mu_{top} > \mu_{all}$	1.82	.61	1.79	.60	1.90	.63	1.92	.64
	H2: $\mu_{top} < \mu_{all}$	0.18	.06	0.21	.07	0.10	.04	0.09	.03
Completeness	Hunc	1.00	.33	1.00	.33	1.00	.33	1.00	.33
	H1: $\mu_{top} > \mu_{all}$	2.01	.67	2.00	.67	1.94	.65	1.97	.66
	H2: $\mu_{top} < \mu_{all}$	0.00	.00	0.00	.00	0.06	.02	0.04	.01
Practicality	Hunc	1.00	.33	1.00	.33	1.00	.33	1.00	.33
	H1: $\mu_{top} > \mu_{all}$	1.85	.62	1.93	.64	1.33	.44	1.22	.41
	H2: $\mu_{top} < \mu_{all}$	0.14	.05	0.07	.03	0.69	.23	0.78	.26

Note: Top-1 N = 124, Top-3 N = 99. BF = Bayes Factor, PMP = Posterior Model Probability. H_{unc} hypothesized no specific direction, H1 for overall creativity (μ_{top}-ori > μ_{all}-ori, μ_{top}-com > μ_{all}-com, μ_{top}-pra > μ_{all}-pra) hypothesized a higher top-score versus the overall score for all three aspects. H1 for originality, completeness and practicality ($\mu_{top} > \mu_{all}$) hypothesized a larger top score than the overall score for this aspect, whereas H2 ($\mu_{top} < \mu_{all}$) a larger overall score than the top score for this aspect.

2.4 Discussion

In line with the CPS model (Isaksen et al., 2011), two CPS tasks for primary education that represent the CPS model were prepared that explicitly embedded the exploration of knowledge (fact finding), defining the problem (problem finding), and the identification of creative ideas (solution finding) alongside the generation of ideas (idea finding). First, we examined whether successful fact finding and problem finding were positively associated with idea finding of primary school students across CPS tasks. Second, we explored whether primary school students were able to select ideas scoring high on multiple aspects of creativity across CPS tasks, or ignored certain CPS aspects.

Question 1.1: Fact Finding & Problem Finding

Regarding the first research question, a successful exploration of knowledge and a high-quality problem definition were positively associated with students' ability to find more and more original ideas. More specifically, the exploration of knowledge seemed especially positively related to the number of ideas a student thought of. This suggests that exploring knowledge on the problem may indeed kick-start the idea finding stage by activating more solution opportunities, as was hypothesized (Isaksen et al., 2011; McCaffrey, 2016). At the same time, we found that a successful exploration of knowledge and a high quality problem definition were not positively associated with students' ability to find complete and practical ideas to tackle the problem. Our results did however suggest that defining the problem helped with finding complete ideas, whereas fact finding might even hamper it. It might be the case that the exploration of knowledge leads a student away from finding complete ideas because it highlights too many opportunities and as such withholds students from focusing on one idea and elaborating on it. Defining the problem might on the other hand assist in finding completing ideas, because it may provide focus in the idea finding stage, as was found by Arreola and Reiter-Palmon (2016). Although more research is needed to further support this hypothesis, discussing the benefits of focus and reflecting on the problem when students do CPS might help them to prepare ideas that are complete. The exploration of knowledge and defining the problem are possibly not helping or even impeding students' ability to find practical ideas because they stimulate students to think outside the box and not to use things that are convenient to use. However, this may differ across CPS tasks, as was indicated by the different results we found across the two CPS tasks we used in our study.

Question 1.2: Solution Finding

With regard to our second research question, results indicated that in general primary school students are able to identify their most creative solutions. We did not find any indications that aspects such as originality, completeness, and practicality were undervalued by the students when defining their top 3. Results did however indicate that the primary school students were much more likely to value rather than disregard completeness. It is possible that students do take all aspects into account but pick the complete ideas first because these ideas took most cognitive effort and time. It could also be the case that students struggle with taking all aspects of solutions into account at once and that completeness is the most salient to them as was found by Rietzschel and colleagues (2010). To gain insight in whether students really appreciate all aspects of creativity or not, a more qualitative approach zooming in on how these solution finding processes take place would be worthwhile.

The influence of tasks

We used two CPS tasks situated in two different domains to explore whether the results could differ across tasks. Although Reiter-Palmon and colleagues (2009) found differences across CPS tasks, our findings tended to be quite in line across the two CPS tasks. Reiter-Palmon and colleagues (2009) did however manipulate the complexity of the tasks, whereas we tried to develop CPS tasks that were different in domain, but comparable in complexity. A small difference between the tasks appeared with regard to how the exploration of knowledge and identification of the problem were associated with the practicality of ideas. It could be the case that due to still existing small differences in complexity one task evoked more practical ideas from students than the other, as is illustrated by the fact that the mean practicality scores differed across the two CPS tasks we used (see Table 1). Future research may provide more insight in the relation between how students perceive the complexity of a task, the application of fact finding and problem finding, and the eventual practicality of ideas.

Limitations

For the top-3 scoring procedure, we had to deal with a relatively small sample size (N = 99). However, the similar findings for the top-1 and top-3 scores suggested this smaller sample size did not negatively influence the results. Furthermore, to keep the CPS task simple we asked students to come up with only one problem statement. In other studies, students were asked to generate multiple problem statements and select one problem statement to work with in the next steps (e.g., Ma, 2009; Reiter-Palmon, 2017). This may have improved the quality of the problem statements and its influence on the idea finding process (cf. Arreola & Reiter-Palmon, 2016). In future studies, it would be interesting to investigate how constructing multiple problem statements impacts idea finding in primary school students. Furthermore, originality, completeness and practicality were rated by external judges. Although students chose ideas that were rated relatively high on creativity as scored by these judges, we do not know why students selected these ideas and whether they completely agree with the scores the judges assigned. A more qualitive, in depth examination about how they evaluate their own ideas would give us more insight into this process. Although this study showed associations between outcomes from different CPS processes, this does not imply causation. Therefore, the results should be interpreted with care.

Practical implications

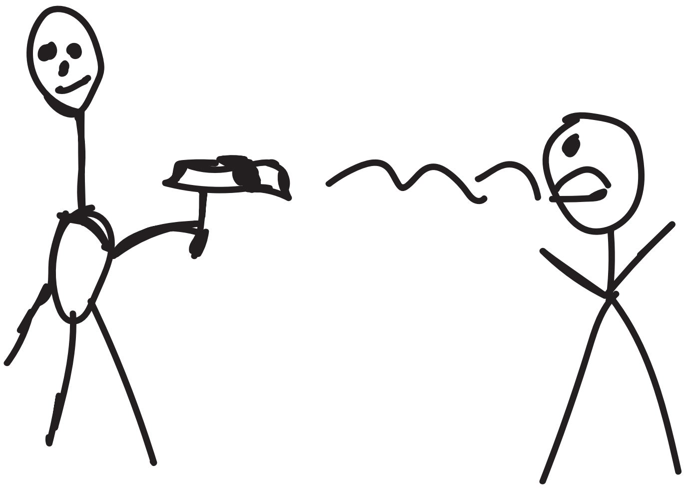
Our findings show that when students struggle to find multiple or original ideas to solve problems, teachers could stimulate students to explore their knowledge on a problem in order to enhance the generation of ideas. When students struggle to find complete ideas, it might be helpful for students to explicitly define the problem at stake. For this study, we developed two CPS tasks for primary education that embedded and scaffolded these processes of fact finding and problem finding, next to idea finding and solution finding processes. Because the CPS stages might not follow this order in practice and students probably switch between stages multiple times while solving a problem (Isaksen et al., 2011; Treffinger et al., 2008), we encouraged students in this study to look back at the first stages while finding ideas. In order for these processes to be transferred to primary education, we think it is important to scaffold the CPS stages first. This might help students to develop a CPS habit that embeds the exploration of knowledge and the identification of the problem before generating ideas. Additionally, explicitly embedding the full CPS process in such a task could help teachers to not only focus on the idea finding stage but pay sufficient attention to the other stages as well (Cropley, 2006; Piffer, 2012). We encourage teachers to take a more flexible approach later on. Additionally, we encourage teachers to discuss the most creative ideas with students. This may give students even more insight in what a final, creative solution could look like.

Conclusions

Our results indicated that when dealing with primary school children and when aiming especially for more and original ideas, it is beneficial to apply fact finding and problem finding before engaging in idea finding. As such, teachers may explicitly pay attention to these preparatory stages. Our study also indicated students are able to identify their most creative ideas and do not undervalue specific aspects of creativity in this solution finding process. In general, findings were rather similar across the two CPS tasks we used. Overall, we think CPS is a welcome addition to the contemporary primary education curriculum, connecting knowledge and creativity in a fun and flexible way.

3

EXAMINING THE ASSESSMENT OF CREATIVITY WITH GENERALIZABILITY THEORY: AN ANALYSIS OF CREATIVE PROBLEM SOLVING ASSESSMENT TASKS



"Put the melting ice pops in a water gun and shoot it in your mouth."

This chapter is based on:
Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2022). Examining the assessment of creativity with generalizability theory: An analysis of creative problem solving assessment tasks. Thinking Skills and Creativity, 43, 100994. https://doi.org/10.1016/j.tsc.2021.100994

Acknowledgement of author contributions:
JvT acquired funding for the study; MvH and TM designed the study; MvH collected the data; MvH and TM planned the data analyses; MvH analysed the data; MvH drafted the manuscript; all authors contributed to critical revision of the manuscript, TM, EK and JvT supervised the study.

Supplementary material can be found at: https://doi.org/10.34894/GIUAOI

Abstract

The assessment of creative problem solving (CPS) is challenging. Elements of an assessment procedure, such as the tasks that are used and the raters who assess those tasks, introduce variation in student scores that do not necessarily reflect actual differences in students' creative problem solving abilities. When creativity researchers evaluate assessment procedures, they often inspect these elements such as tasks and raters separately. We show the use of Generalizability Theory allows researchers to investigate CPS assessments in a comprehensive and integrated way. In this paper, we first introduce this statistical framework and the choices creativity researchers need to make before applying Generalizability Theory to their data. Then, Generalizability Theory is applied in an analysis of a CPS assessment procedure. We highlight how alterations in the nature of the assessment procedure, such as changing the number of tasks or raters, may affect the quality of CPS scores. Besides this, we present implications for the assessment of CPS and for creativity research in general.

3.1 Introduction

The assessment of creativity is a challenge (Cropley, 2000; Piffer, 2012; Sternberg, 2020). Researchers develop creativity assessment procedures to measure creativity, for instance to determine students' creative abilities in educational settings. A creativity assessment procedure includes the full range of chosen methods for evaluating a student's creative performance (Gipps, 1994). This could for instance mean a variation of tasks, different raters or rating procedures (e.g., Bouwer et al., 2015; Crossley et al., 2011). However, these elements of an assessment procedure usually also introduce variation in scores which does not necessarily reflect actual differences in students' creativity. Specifically for the assessment of creativity, domain-specific creative abilities, creativity task characteristics, rater biases in, for example, originality scores, and the chosen rating procedures are all examples of potential sources of variability in creativity scores, which are not due to actual differences between students' creativity (Amabile & Pillemer, 2012; Conti et al., 1996; Guo et al., 2019; Kaufman et al., 2008; Reiter-Palmon et al., 2009). For creative problem solving (CPS) in particular, determining these potential sources of variability might be of interest. CPS is a measure of little-c, that is, everyday creative achievement, in which students use subject knowledge and creativity to solve everyday problems creatively (Craft, 2011; Isaksen & Treffinger, 2004; Okuda et al., 1991). The problems students solve may stem from different domains and students may as such have different subject knowledge or previous experiences with the problem, impacting the generalizability of scores on CPS tasks (Reiter-Palmon et al., 2009). Raters on the other hand might have more affinity with a certain problem or might have preliminary ideas about what solutions could or could not work (Long, 2014). Again, this may impact CPS ratings and consequently the generalizability of students' CPS scores. In divergent thinking studies, ideas are most often assessed on fluency, flexibility, and originality (Reiter-Palmon et al., 2019). In open-ended tasks like CPS tasks, raters often apply more criteria such as how appropriate (or practical) and thoughtful (or complete) ideas are (Long, 2014). Reviewing how generalizable CPS scores for these criteria are, may give us insight in specific biases in CPS assessments and how these differ from biases in divergent thinking tasks (e.g., Hass et al., 2018; Silvia et al., 2008). In addition, knowing to what extent elements of a CPS assessment procedure and their interactions cause variance in addition to actual differences in students' CPS ability helps interpreting students' scores and points to elements of the CPS assessment procedure that potentially require improvement.

Besides determining the influence of the separate elements of the CPS assessment procedure, assessment developers also want to determine the quality of the assessment procedure as a whole. Calculating to what degree the assessment procedure produces reliable scores for the constructs of interest assists in determining this psychometric quality. When there is a lot of "noise" in student scores, adaptations in the assessment procedure, such as more problem tasks or more raters, might be needed to decrease its impact on the reliability of the results. However, more problem tasks and raters do also implicate using more time and resources. Especially when the aim is to embed the assessment of CPS in the curriculum, procedures need the be as efficient as possible. Thus, neither too many nor too few problem tasks and raters should be used. Generalizability Theory (Brennan, 2001) provides a statistical framework which assists in making these informed decisions about assessment procedures. Several authors presented a general overview of this theory and how it can be applied (e.g., Brennan, 2010; Mushquash & O'Connor, 2006). In this paper, we want to complement this work by offering a non-technical manual specifically in the CPS context. After introducing Generalizability Theory, including an explanation of the decisions researchers need to make, we will demonstrate how Generalizability Theory can be applied on CPS data with a step-by-step manual. By providing information on the statistical formulae, the data structure, the syntax, and the output, we invite creativity researchers to try Generalizability Theory themselves and experience how it may enhance their development and evaluation of creativity assessment procedures and CPS assessments in particular.

Generalizability Theory

Generalizability Theory provides a statistical framework for investigating and developing reliable assessment procedures. Generalizability is regarded as a form of reliability or reproducibility that is important for estimating to what extent observations can be used to make claims about a student's true ability (Brennan, 2010). In classical test theory, as opposed to Generalizability Theory, measurement error caused by multiple sources is regarded as random and intertwined; Classical test theory does not allow for one to disentangle the sources of error. Generalizability Theory, however, determines the sources of error variation, disentangles them, and estimates the error for each source. By reviewing this error, Generalizability Theory inspects how consistent elements of an assessment procedure, such as tasks and raters, 'behave' and transforms this behavior into a coefficient. These coefficients serve three purposes: (1) they can be reviewed to determine the impact of one of the elements in an assessment procedure, (2) they can be used to calculate the generalizability of scores produced by the assessment procedure as a whole, and (3) they can be used to calculate how potential changes in the assessment procedure would impact the generalizability of the scores. Generalizability Theory proposes two stages of analysis: the Generalizability study (G-study) and the Decision study (D-study). Where the G-study estimates the error attributed to the elements in the assessment procedure and calculates current generalizability, the D-study uses the estimated errors to predict how alterations in the assessment design (e.g., more or less tasks or raters) would influence generalizability.

G-study

In the G-study stage, error (i.e., variability in student scores which does not originate from differences in student creative ability) attributed to the separate elements of the assessment procedure, is identified and the relative importance of all elements is estimated. The magnitude of the variance each aspect (and the interaction of aspects) introduced to the test score is reflected in a variance component. The amount of variance that can be rightly attributed to students' true ability can be reviewed here by reviewing the variance component of the students. Besides this, the impact of the other elements, such as tasks, raters or rating procedures and their interactions, can be reviewed and compared as well. Within a G-study, the generalizability of a certain score (e.g., a creativity score) is described in a coefficient that is calculated based on the variance components. This generalizability coefficient reflects the accuracy of generalizations made from the observed scores compared with the universe score. A universe score is the expected value of a student's observed score over all tasks to which an assessment procedure wants to generalize. By reviewing this generalizability coefficient, the researcher can gain insight in the psychometric qualities of the procedure. Often, a generalizability of .70 or higher is regarded as acceptable when we aim to measure abilities such as creativity (Brennan, 2010).

D-study

In the second stage, multiple Decision-studies (D-studies) can be carried out. The ultimate goal of a D-study is to increase the psychometric quality and efficiency of an assessment procedure by choosing the optimal number of elements (e.g., tasks and raters) for an assessment procedure to produce generalizable results in a specific context. This can be done by calculating the generalizability for hypothetical numbers of elements using the variance components obtained in a G-study (e.g., Hass et al., 2018). In this way, it is explained how alterations in the assessment procedure (e.g., reducing or extending the number of creativity tasks involved) will affect generalizability. Multiple D-studies can be used to estimate the number of tasks and/or raters that is needed for a desired level of generalizability.

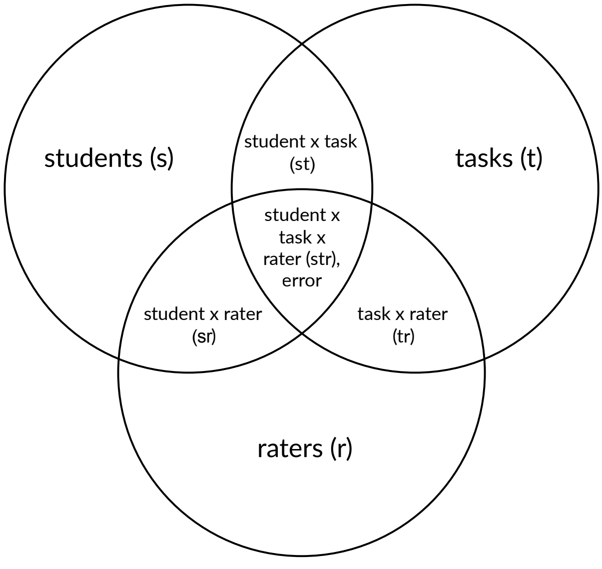
Choices to consider

Before applying Generalizability Theory, researchers have some choices to make. First they need to determine what the universe is they aim to generalize to (Brennan, 2010). In other words, researchers need to describe what specific ability or trait they aim to measure. Especially in light of the discussion on the task or domain specificity of creativity, this is an important aspect to consider (e.g., Baer, 2012; Barbot, et al., 2016b; Reiter-Palmon et al., 2009). Researchers could for instance decide to generalize to a specific creative process across domains (e.g., problem construction or idea selection; Arreola & Reiter-Palmon, 2016; Zhu et al., 2020) or to a creative ability within a more specific domain such as musical or mathematical creativity (Barbot & Lubart, 2012; Schoevers et al., 2020a). The chosen universe determines the focus of the assessment procedure and the nature of the elements involved.

Next, researchers need to decide what elements of the assessment procedure to include into their G-study. In Generalizability Theory, the elements of the assessment procedure (such as tasks, raters, or occasions) are usually called facets. Facets are similar to factors used in an Analysis of Variance (ANOVA; Brennan, 2010). Figure 1 shows a conceptual model of possible facets of a creativity assessment procedure and the interactions between these facets.

Figure 1

Ven diagram of the facets of an assessment procedure and their interactions.



Facets may for instance include tasks, domains, raters, or measurement moments. The researcher selects those facets that may impact test scores the most. For instance, what rating procedure is used may impact scores more than test settings such as whether the tasks are administrated in or outside of class. Here, it is also important to consider the aspect of time and resources. Researchers need to consider issues like how many tasks would be feasible in a given timeframe and how many raters can be involved, both now and in future assessment. In an analysis based on Generalizability Theory, researchers can choose to treat facets as fixed or random. Cronbach and colleagues (1972) suggest to apply a fixed facet when (1) the test developer deliberately includes certain facets and is not interested in generalizing beyond them or beyond the observed levels or (2) when the number of potential levels or variations of a factor is that small that all levels are already included in assessment procedure. In other words, when facets are treated as fixed, the derived inferences are limited to the facets in an assessment procedure, such as specific tasks or types of scoring (Silvia et al., 2008). Researchers may treat facets as random if they, for instance, aim to apply many similarly qualified raters or when they regard the tasks as a selection of many possible tasks. This may for instance be the case when (trained) teachers assign scores in the assessment procedure or when classroom projects are rated to obtain creativity scores. In other words, facets are regarded as random when there is a theoretically infinite pool of tasks and raters. If we would randomly select a set of tasks or raters, they should produce roughly the same observed scores as another randomly selected set (Silvia, et al., 2008).

Besides deciding what facets to include, researchers ideally also decide beforehand if they want to make absolute or relative decisions in the assessment procedure. For absolute decisions, such as assigning IQ scores, the aim is to index the individual student's absolute level of ability, independent of the performance of others. Test developers usually aim for absolute decisions when they develop, for example, a standardized creativity test. A student's performance has improved when this student gets a higher absolute score based on the results of the creativity test, compared with the score this student got based on the results of an earlier test. When researchers decide to make relative decisions, they are merely interested in how a student is positioned compared with other members of a group. These creativity tests aim to measure, for example, whether a student has become more creative over time compared with the rest of this student's class. This may for instance mean a student receives a lower absolute rating on a second test, but because on average the group performed even worse, the student's performance might still have improved. Generalizability Theory provides generalizability coefficients for absolute and relative decisions separately and calculates these with the use of the variance components.

When researchers aim to make absolute decisions, all variance components, including the main variance components of the raters, tasks, and the rater-task interaction, are included in the calculation of the absolute generalizability (see formula 1). This makes the absolute D-coefficient more stringent than the relative g coefficient. With the relative g coefficient, we are interested in the relative standing of students. As such, the g-coefficient only considers variance components (partially) attributed to the students (formula 2). Because the absolute ratings and scores on tasks do not matter here, the main variance components and the interaction of the tasks and raters are not included in the calculation of the relative g coefficient. The formulae below need to be adopted to the specific assessment design (e.g., Figure 1). For instance, when besides tasks and raters different rating procedures (p) are applied, this is added to the equation as a fourth facet (e.g., sp/n for relative generalizability).

1. $$\text{Absolute generalizability (D)} = \frac{s}{s+t/n+r/N+st/n+sr/n+rt/n+str,\ e/n}$$

2. $$\text{Relative generalizability (g)} = \frac{s}{s+st/n+sr/n+str,\ e/n}$$

where
s = variance component of students
t = variance component of tasks (not included when making relative decisions)
r = variance component of raters (not included when making relative decisions)
st = variance component of student x task
sr = variance component of student x rater
rt = variance component of rater x task (not included when making relative decisions)
str,e = student x rater x task interaction and error (confounded)
n = number of tasks and/or raters.

3

As a last decision, researchers need to determine what software to use to perform the statistical analyses. Programs that allow researchers to apply Generalizability Theory are inter alia IBM SPSS Statistics, SAS and RStudio (Jiang, 2018; Mushquash & O'Connor, 2006). Because our analysis of CPS tasks includes a relatively straightforward statistical model, IBM SPSS Statistics was used to execute the Generalizability study. SPSS is a widely used program to manage data and the provided syntax (see online materials) allows developers of creativity assessments to quickly perform a generalizability study, for instance in the piloting phases of their research. Especially for more complex models such as multivariate models, RStudio is considered to be a suitable alternative (e.g., Jiang et al., 2020).

Generalizability Theory in creativity research

Most researchers investigated the elements from a creativity assessment procedure separately (e.g., Benedek et al., 2016; Ceh et al., 2022; Reiter-Palmon et al., 2009), for instance, by focusing on what kind of tasks should be included or how many raters would be necessary to obtain a reliable indication of one's creativity. However, in such research, it is not taken into account how one element (e.g., raters) may interact with other elements (e.g., tasks or students) and how this affects the generalizability of creativity scores. In a variety of educational domains other than creativity, researchers have used Generalizability Theory to investigate assessment procedures. For instance, Bouwer and colleagues (2015) added genre as a facet and studied the generalizability of writing scores. Crossley and colleagues (2011) and Moonen-van Loon and colleagues (2013) compared different workplace-based assessment scales to determine which scales produces generalizable results in a limited number of assessments. Hill and colleagues (2012) applied Generalizability Theory to data generated with a teaching observation tool to illustrate how one can determine the number of raters or lessons that need to be included to determine a teacher's teaching quality.

In creativity research, only a few researchers have applied Generalizability Theory to their data. Or as Myszkowski and Storme (2019) put it: "in creativity research Generalizability Theory is certainly underused, as the decisions relative to the examination of reliability are rarely put in relation with the intended interpretation of the score" (p. 169). The few studies that were published mainly focused on divergent thinking scoring procedures. For instance, Silvia and colleagues (2008) asked undergraduates to complete three types of divergent thinking tasks: alternative uses items (e.g., "think of alternative uses of a brick"), instances items (e.g., "think of unusual things that are round"), and consequences items (e.g., "think of consequences when human no longer need sleep"). With the use of Generalizability Theory, they compared average scoring (i.e., every single idea is evaluated separately, scores are averaged across ideas) with Top 2 scoring (i.e., only two ideas selected by the participant are scored). The results implied that, for both methods, only two or three raters rating a single alternative uses or instances item were required to obtain reliable scores to make relative decisions. For the consequences item, four or five raters would be needed. In a second study, Hass and colleagues (2018) applied layperson ratings on the same dataset and compared two types of scoring. The authors concluded that with average scoring three layperson raters and four alternative uses items could produce reliable scores to make relative decisions. For the consequences item, three layperson raters and eight items would be needed. For both items, the alternative snapshot scoring system (i.e. assigning a single score to the entire set of responses generated) could not produce reliable scores, even with an increase of raters.

In the domain of creative writing, Kaufman and colleagues (2007) applied the Consensual Assessment Technique (cf. Amabile, 1996) and concluded 5 raters and 15 captions of every student would be needed to reliably assess students' caption writing ability and make absolute decisions. Long and Pang (2015) focused on science and studied the generalizability of student, teacher, and researcher creativity ratings of two open-ended tasks. They concluded (too) many raters were necessary to obtain reliable results for both relative and absolute decisions. Again, the rater inconsistencies tended to be different across tasks.

The current paper adds to the existing generalizability studies by illustrating how Generalizability Theory may impact the development of a CPS assessment procedure. For CPS in particular, little research is conducted on the generalizability of scores, especially in a young age group. Within this study, primary school students are asked to diverge and give multiple solutions to each problem (cf. Reiter-Palmon & Arreola, 2015) and a form of average scoring is applied. This allows us to compare findings of this study with the generalizability studies on divergent thinking data. Alongside the generalizability of originality scores, completeness and practicality scores will be evaluated as well in this study. Reviewing how generalizable CPS scores for these criteria are, may give us insight in specific biases in CPS assessments and how these differ from biases in divergent thinking tasks. Therefore, we conducted a G-study and multiple D-studies with a CPS assessment dataset. To encourage creativity researchers to consider this method too, we describe the study step by step alongside information on the data structure, the syntax, and the output.

3.2 An Analysis of Creative Problem Solving Assessment Tasks

To face the unknown problems of tomorrow, schools need to foster CPS abilities in their students from an early age on (Craft, 2011). Teachers however, struggle with teaching CPS and assessing students' CPS abilities (Kettler et al., 2018). Within this CPS assessment project, we are developing tasks and rating procedures to measure CPS in primary education. Three G-studies and a series of D-studies were carried out with CPS assessment data from this project. The data, syntax and output are available for download via https://doi.org/10.34894/GIUAOI.

Choices made

The 'universe' we aimed to generalize to is upper primary school children's CPS abilities. This means we aim to measure children's everyday abilities to solve problems creatively across domains. The problem tasks we included in our assessment procedure could therefore be regarded as a selection of many possible problems that children can encounter in daily life. As such, we included the tasks in our generalizability study as a random facet. Eventually we would like teachers to rate the CPS tasks themselves. Therefore, we wanted the raters to be regarded as interchangeable and as such included the raters as a random facet too. Because the students included in the sample were a selection of many possible students, they were regarded as a random facet as well. For present purposes, we evaluated CPS assessment scores in the light of both absolute and relative decisions.

Study design

Six classes of 4th and 5th grade students (n = 137; mean age = 10.50; 53% girls) from three schools participated voluntarily in the study (see also Chapter 2 for more information on the tasks and procedure). Two problem tasks were completed by all students. One problem from the science domain and one from the social domain were selected in order to include distinctive creativity domains (Kaufman, 2012; Kaufman et al., 2009). The science problem described two children, Lisa and Tom, buying ice cream in the supermarket. The problem they encountered was that the ice cream melted on the way home. The social problem described Simon, who gets distracted in class by his friend Julian all the time. The problem Simon needed to solve was how to stop his friend's behavior without this having a negative impact on their friendship. The problems were presented written on paper and were read to the students by the researcher.

After going through two preparatory steps in which students defined their prior knowledge and described the problem at stake, students received 10 minutes to list as many different and original ideas as they could that would solve the problem. All the ideas were rated on originality, completeness, and practicality by two raters (one graduate and one post-graduate), using a modified version of the Consensual Assessment Technique with a 5-point Likert scale (e.g., Amabile, 1996; Byrne et al., 2010; Corazza, 2016; Okuda et al., 1991; Reiter-Palmon et al., 2009).

Table 1
Descriptives for CPS Scores of Two Raters on Two Tasks (N = 137).

	Science Problem Task				**Social Problem Task**			
	M	**SD**	**Min**	**Max**	**M**	**SD**	**Min**	**Max**
Originality								
Rater 1	0.74	0.54	0	2.67	0.51	0.57	0	2.83
Rater 2	0.38	0.38	0	2.00	0.26	0.40	0	2.50
Completeness								
Rater 1	2.48	0.51	1.20	4.00	2.44	0.40	1.33	3.71
Rater 2	2.36	0.53	0.75	4.00	2.18	0.43	0.67	3.40
Practicality								
Rater 1	2.75	0.65	0.80	4.00	3.36	0.60	1.28	4.00
Rater 2	2.66	0.54	0.80	3.57	3.11	0.59	0.83	4.00

Preparing the data

For every rater, the scores for originality, completeness and practicality were averaged for the separate tasks (Table 1). Next, the data was reshaped from a wide format (all student's scores in a single row, with each rating for every task in a separate column) to a long format (each row includes one rating on one task; see online materials).

Execution of de G-study: Preparing the syntax

In the first stage of our analysis, three G-studies were conducted (i.e., for originality, completeness, and practicality) to determine the variance components attributable to the different facets (students, tasks, raters and their two-way interactions; Figure 1) for every CPS aspect. The G-studies were executed using syntax in SPSS version 24. The syntax for the G-study is provided in the online materials.

We aimed to decompose the variance of the students, the rater, and the task. Therefore, the first line of the syntax included the VARCOMP command, followed by the dependent variable (originality, completeness, or practicality) and the BY statement followed by our facets: the students, the tasks and the raters. Any covariates authors wish to control for can be included in this first line as well by adding a WITH statement followed by the desired covariate. Because the students, tasks and raters in this study can be regarded as a selection from many possible students, tasks and raters, we included these facets in our analyses as random with the /RANDOM subcommand. By including a /FIXED line in the syntax, researchers may include facets that are not interchangeable and regarded as fixed such as a specific scoring system. The /DESIGN subcommand describes the facets (and their interactions) of which the variance components should be modelled. Because the variance components of both the main facets and their interactions could provide us with relevant information, all facets and their interactions were included in our syntax. In our design, potential three-way interactions (student x task x rater) were confounded with the residual error. Any nested effects in the design can be specified after the /DESIGN command as well using a WITHIN statement (e.g., task WITHIN domain; cf. Mushquash & O'Connor, 2006 for more complex nested models). The /METHOD command includes the estimation method. We choose to apply restricted maximum likelihood (REML) because the maximum likelihood (ML) estimator might be negatively biased with smaller sample sizes, resulting in smaller variance components. For moderate deviations from normality, researchers can choose the MINQUE method as well here. Because we needed an intercept to calculate how our facets induce variance around this intercept, it was included with the /INTERCEPT = INCLUDE command.

3

Reviewing the output

By running the syntax, three tables were produced as output (see the online materials). The first table included factor level information. This allowed us to inspect how many scores were included per student (4), per rater (2 * 137 = 274) and per task (2 * 137 = 274). Any missing data could be reviewed in this table as well. The second table included the main outcomes from the G-study, i.e., the variance estimates for every facet and the interactions. The last table included the covariance matrix that was used to calculate these estimates. We ran the syntax separately for the originality, completeness and practicality scores, collected the variance components from the second table and transformed them into percentages of the total variance (Table 2).

Table 2

The Results of the G-study. Variation (s^2) Attributed to the Facets (p, t, r) and their Interactions (st, sr, rt, str,e).

	Originality		Completeness		Practicality	
	s^2	%	s^2	%	s^2	%
Student (p)	0.064	22.22	0.053	21.81	0.141	27.54
Task (t)	0.013	4.51	0.004	1.64	0.137	26.76
Rater (r)	0.044	15.28	0.015	6.17	0.011	2.14
Student*Task (st)	0.100	34.72	0.052	21.40	0.147	28.71
Student*Rater (sr)	0.002	0.69	0.000	0.00	0.000	0.00
Rater*Task (rt)	0.002	0.69	0.004	1.65	0.005	0.98
Student*Task* Rater / error (str, e)	0.063	21.89	0.115	47.33	0.071	13.87
Total s^2	0.288	100.00	0.243	100.00	0.512	100.00

Interpretation of the G-study

By reviewing the variance of the student facet, the amount of variance that can be rightly attributed to students' true ability was examined first. For all three aspects of CPS, somewhat similar amounts of true student score variance were found. For originality and completeness, about 22% of the variance could be attributed to actual differences between students. For practicality about 28% could be attributed to students. This means that despite the effort to focus on students' true ability, a large share of 72 to 78% of the variance was attributed to other sources. Next, the variance components of the task and rater facet were reviewed to get an idea of how these facets impacted the CPS scores. For originality and completeness, the task itself did not seem to introduce a lot of variability (only 4.51% and 1.64%). The much higher percentage of variance due to the task for practicality (about 27%) implies that the influence of our tasks on the CPS scores for practicality was almost as strong as the influence of the students who completed these tasks. This is also apparent in the average differences in practicality scores in Table 1. Raters did not differ that much in their scoring of completeness and practicality, illustrated by the 6% and 2% variance at the rater level. For originality, the variance at the rater level was about 15%. Thus, raters seemed to disagree more in their originality ratings than in their completeness and practicality ratings.

When it comes to the interactions between facets, the student x task interaction of all three CPS aspects accounted for a large part of the variance (about 34% for originality, 21% for completeness, and 29% for practicality). This implies that for the students' CPS scores, it mattered which student completed which tasks. This might for example be due to task-specific or domain-specific CPS abilities. Almost no variance (max 0.69%) was found on the student x rater level, which implies that raters were not biased towards certain students. Because the students were unknown to the raters, this was as expected. There was also very little (0.69% to 1.65%) variance on the rater x task level, which indicated raters did not differ in their judgements depending on the task they rated. This indicated that our rating scheme may be used across various problem tasks. Additionally, there were considerable amounts of variance left at the student x task x rater and error level. Especially for completeness, this percentage of residual variance was quite large (about 47%).

Next, the generalizability of the applied assessment procedure was reviewed. The formula (Formula 1 & 2) was used to calculate the absolute (D) and relative (G) generalizability that reflect the generalizability for the set of tasks and raters we used (two tasks, two raters; Table 3).

Table 3
Relative and Absolute Error Variances and Generalizability for Two Tasks and Two Raters.

	Originality	Completeness	Practicality
Absolute error variance	0.29	0.24	0.51
Relative error variance	0.17	0.17	0.22
Absolute generalizability (D)	.40	.45	.46
Relative generalizability (g)	.49	.49	.61

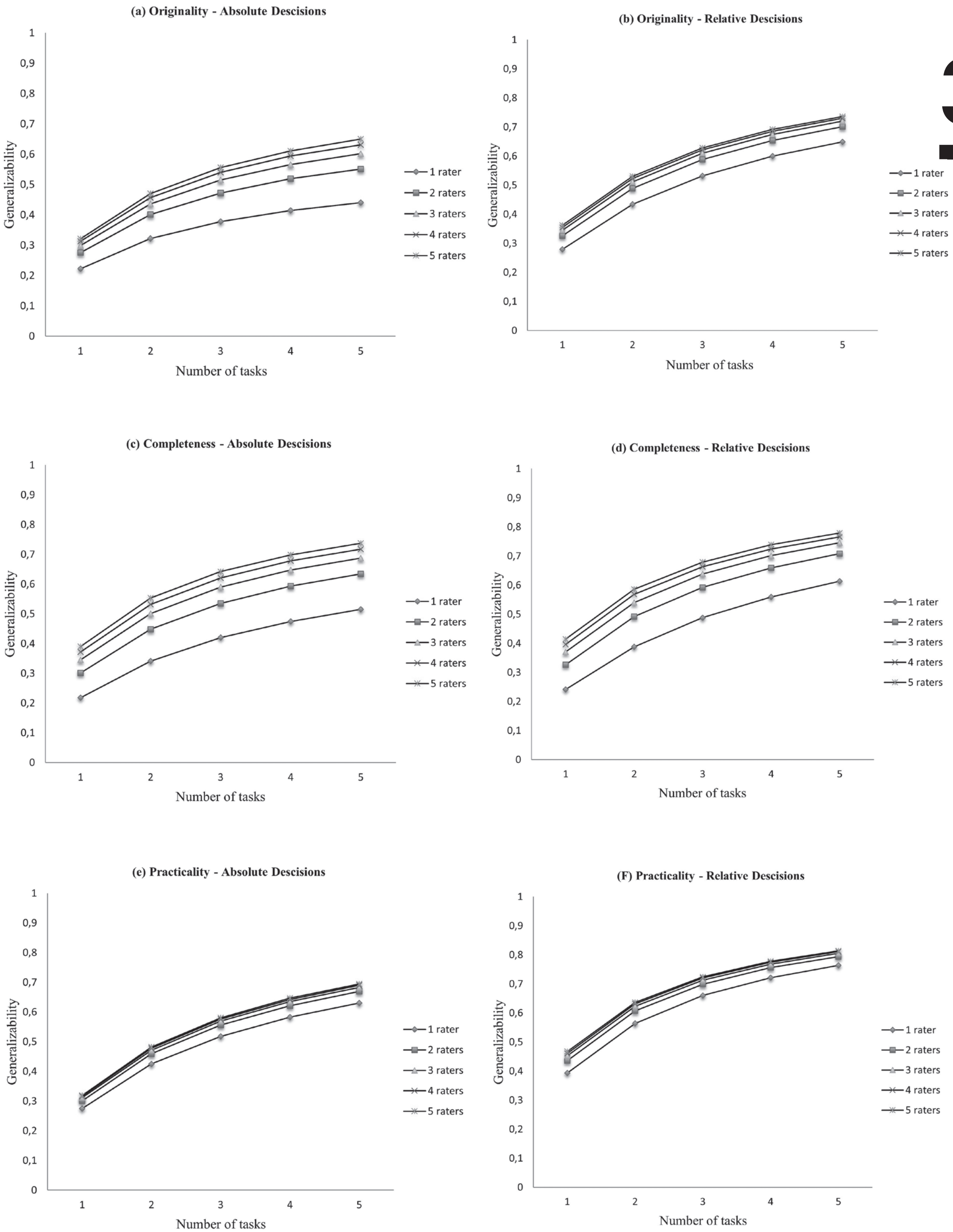
Both the absolute (D = .40 - .46; Table 3) and relative (g = .49 - .61; Table 3) coefficients did not reach the desired level of .70 (Brennan, 2010). Therefore, the current assessment procedure including two tasks and two raters could be regarded as insufficient to make both absolute and relative decisions. Therefore, multiple D-studies were conducted to get an idea of how many tasks and raters should be included in the assessment procedure to reach the desired generalizability level of .70 for originality, completeness, and practicality.

Execution of the D-Study

Multiple D-studies were executed for the three CPS aspects originality, completeness, and practicality and for both absolute (norm-based) and relative decisions. The numbers of tasks and raters in our assessment design were varied and generalizability was recalculated using the percentages of the variance and the formulae (Table 2; Formula 1 & 2). For instance for originality, five tasks and two raters would produce an absolute generalizability of about .55 (Absolute Generalizability (D) = 22.22 / (22.22 + 4.51/5 + 15.28/2 + 34.72/5 + 0.69/2 + 0.69/(10) + 21.88/(10)) and a relative generalizability of about .70 (Relative generalizability (g) = 22.22/(22.22 + 34.72/5 + 0.69/2 + 21.88/10). The number of tasks and/or raters was varied until an absolute and relative generalizability of at least .70 was reached.

Figure 3

Graphical representation of the results of the D-studies. The X-axis represents the number of tasks, the Y-axis the generalizability of the scores. The separate lines represent the number of raters.



Interpretation of the D-Study

According to these predictions, for absolute decisions about a student's ability to produce original ideas, at least ten CPS tasks rated by four raters are necessary to reach a generalizability of .70. When aiming for only two raters, at least thirty tasks should be used. For completeness, four tasks and four raters should be used to make absolute decisions, and for practicality eight tasks and two raters. (Figure 3). Because for relative decisions the absolute ratings on tasks do not matter, a lower number of tasks and raters needs to be included in the assessment procedure. For instance, only five tasks rated by two raters are necessary to provide generalizable originality and completeness ratings. For practicality, this would be even less: only three tasks rated by two raters. In sum, the number of tasks needs to be increased to make absolute and relative discussions for all aspects. Increasing the number of raters would positively influence the absolute generalizability of the originality and completeness ratings. With these results in mind, more CPS tasks can be developed, and rating procedures can be refined, to enhance the CPS assessment procedure.

3.4 Discussion

Within this report, we aimed to show how creativity researchers can use Generalizability Theory to make informed decisions about creativity assessment procedures and CPS assessments in particular. The G-study on the CPS data not only provided us with valuable information on where to start with improving the CPS assessment system but also gave us insight in the assessment of CPS.

Implications for the assessment of CPS

The assessment system under investigation in this study consisted of two CPS tasks and two raters. If only this set of tasks and raters would be used to assess CPS abilities, students' scores would reflect other factors than solely student CPS performance and, in other words, would be biased. The earlier generalizability studies on divergent thinking tasks showed that when scores are averaged across the generated ideas a small increase of the number of tasks (e.g., 4 to 8; Hass et al., 2018) and raters (e.g., 2 or 3; Silvia et al., 2008) may suffice to obtain reliable scores to make relative decisions. Within our CPS assessment procedure, students also generated multiple solutions to a problem and a similar way of scoring was applied. The series of D-studies indicated that also for the assessment of CPS, three to five tasks rated by two raters may be applied to obtain reliable CPS scores to make relative decisions. This suggests that when an idea generation stage is applied, average scoring works for assessments of CPS as well. For educational practice, this means that a reasonable number of tasks (e.g., five tasks throughout an academic year) and raters (e.g., two teachers) may suffice to get an image of students' CPS abilities. However, more research is necessary to conclude that teacher ratings of students' CPS abilities are aligned with trained rater assessments of these abilities.

Previous studies on divergent thinking showed that it tends to be difficult to get raters in line on what ideas are original and what ideas are not (e.g., Benedek et al., 2016; Grohman et al., 2006; Guo et al., 2019; Silvia et al., 2008). The seemingly large impact of the rater on the originality scores compared to the impact on completeness and practicality scores suggests that for CPS this is the case as well. The applied rating procedure for originality could be reviewed again to minimalize these disagreements in future CPS assessments in education.

This study also focused on ratings of completeness and practicality, uniquely for CPS. For completeness, a relatively large part of the variance could not be attributed to the students, tasks, and raters. This might be explained by a three-way student x rater x task interaction. It may be that for specific students, the raters scored completeness differently within tasks. This may also mean that a different factor that should be controlled for influenced the scores for completeness, such as perceived problem complexity. More research is necessary to explore both hypotheses. The large share of the variance attributed to the tasks for the practicality scores indicates that the ability to come up with practical ideas when doing CPS highly depends on the type of problem situation or task presented to the students. Differences across tasks were found in many creativity studies (e.g., Hass et al., 2018; Long & Pang, 2015; Reiter-Palmon et al., 2009; Silvia et al., 2008). For the CPS assessment procedure, this means that we might need to include more tasks to assess practicality. However, the results of the D-studies also indicated that for the different indices of CPS, the impact of increasing the number of tasks and raters may vary. For originality and completeness for instance, it seems more worthwhile to increase the number of raters, especially if we intend to make absolute decisions. This does not have to be the case for every creativity assessment procedure using the consensual assessment technique. In other assessment procedures, the number of tasks or raters may even be reduced (e.g., Kaufman et al., 2007). Executing D-studies may help creativity test developers in gaining sufficient data but also helps to prevent administering unnecessary tasks.

In this Generalizability study, the ‘universe’ we aimed to generalize to was upper primary school children’s CPS abilities. In other words, we studied CPS as an ability that we aimed to generalize across tasks or domains. Although more research is necessary to validate the relatively new tasks (see also Chapter 2), the high variance on the student x task interaction may already indicate that CPS – like many other forms of creativity- is a task-specific or domain-specific ability (Baer, 2012; Barbot et al., 2016b; Reiter-Palmon et al., 2009).

Implications for creativity research

Generalizability Theory could be of assistance to study specific issues in the creativity literature, such as the domain-specificity of creative abilities. By including multiple tasks within and across domains, researchers could study the variance within and between domains. As Reiter-Palmon and colleagues (2009) suggested, tasks characteristics such as task complexity, involvement, and forms of self-efficacy (e.g., whether a student is convinced he or she can solve the problem) may influence creative outcomes even within domains. Such task characteristics may cause variance in student scores as well and may as such be included in the G-study.

As noted earlier, more elements of a creativity assessment procedure can be added as a facet in a G-study. For instance, the type of rater (e.g., trained expert raters versus teachers or layman raters; Hass et al., 2018; Kaufman et al., 2013) or the type of rating procedure (e.g., snapshot scoring, Long & Pang, 2015; Silvia et al., 2008) could be included in the analysis to see how they affect the generalizability of the results. Creativity researchers working with Generalizability Theory should however carefully select the facets they include in their design, because they may miss elements that (strongly) impact test scores. Researchers should be aware that they could overlook something that is also called a “hidden facet” (Shavelson & Webb, 1991). A hidden facet can for example be revealed by two facets that cause error in similar ways: when the levels of one facet change, so do the levels from another. This may for instance happen when tasks are administrated in the same order every time. Here, the variance found between the different task scores might be interpreted as task variance, while in fact it could also be an order or learning effect. It is therefore important to consider possible hidden facets in the assessment design.

Further, Generalizability Theory relies on increasing (or decreasing) the number of observations of facets with high (or low) variance (e.g., adding more tasks and raters) to reach the desired level of generalizability. This could however result in unpractical designs. This is illustrated with our analysis of CPS tasks: with only two raters at least 30 tasks are estimated to be necessary to make absolute decisions about a students' ability to produce original ideas. As mentioned before, creativity test developers should therefore ideally define their aims beforehand (Myszkowski & Storme, 2019). Do they aim to make absolute decisions and develop a standardized creativity test, or do they want to take a more relative approach? Often, creativity test developers may be merely interested in a student's relative performance. Also, as some argue, it is almost impossible to develop a standardized creativity test for education (Harris, 2016). Either way, as Silvia and colleagues (2008) state, it would be valuable to do a Generalizability study in the piloting stage of the development of an assessment. This will give researchers insight in the sources of error and could assist in determining how many facets such as tasks and rater would be necessary to generalize the scores to a specific ability.

Within this report, we discussed in detail how Generalizability Theory may be applied in creativity studies. The information on the analysis including the data structure, the syntax, and the output may help creativity researchers to apply this theory to their data. We choose this method because it is relatively easy to apply and interpretation is not that difficult for researchers familiar with ML methods like ANOVA or linear regression analysis. Besides this, Generalizability Theory is very useful in providing information on overall decisions about the design of an assessment procedure (Linacre, 1996). The calculation of both absolute and relative generalizability and subsequent D-studies assist in making these decisions. Researchers interested in more specific information at the individual (e.g., student) level might be interested in Many Facet Rasch Modeling (MFRM; Linacre, 1996; Primi et al., 2019). Although studies applying both techniques in general report a comparable impact of facets (e.g., Smith & Kulikowich, 2004; Sudweeks et al., 2004), MFRM can provide specific suggestions regarding the characteristics of the added facets, e.g., like adding more lenient raters or more difficult tasks (Linacre, 1996).

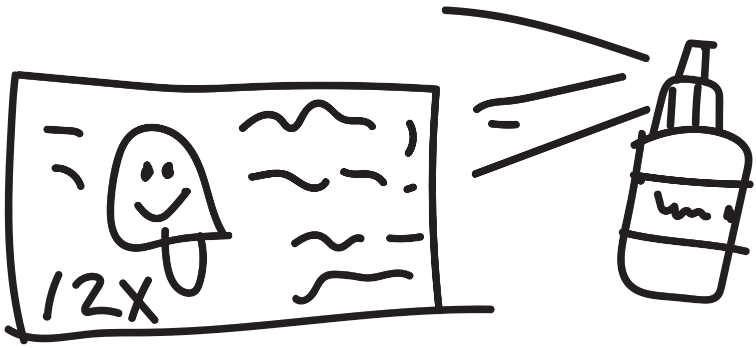
Limitations of the theory

As with every theory, Generalizability Theory comes with some limitations (Brennan, 2010). In balanced designs (when every student completes the same number of tasks and receives the same number of ratings), the calculation of the variance components is straightforward and the estimations are mostly unbiased. When tasks are rated by a different number of raters or when every student completed a different number of tasks, an unbalanced design occurs. Here, the estimation of variance components is more complex and a G-study might produce biased variance components. Consequently, biased D-study results might follow. Brennan (2001) describes G-study and D-study solutions for some frequently encountered unbalanced designs. In the case of an unbalanced design, we recommend researchers to consult this work.

Researchers should also be aware that variances are calculated based on the facets that are included in the design. Adding more facets alters the estimation of all other facets as well. In addition to that, the generalizability of an assessment procedure should be reanalyzed and recalculated when the assessment design is altered to make sure adjustments were indeed an improvement. Also, each creativity assessment procedure needs its own G-study and D-study, as it is likely to have its own set of unique tasks and rating procedures. Findings from one generalizability study cannot simply be applied to another assessment procedure. Nevertheless, using more G-studies and D-studies in creativity research may not only enhance the critical evaluation of creativity tests, it may also help to move towards including the assessment of creativity and CPS in particular more efficiently in everyday educational practice.

4

CREATIVE PROBLEM SOLVING IN PRIMARY SCHOOL STUDENTS



"Keep the ice pops frozen with newly invented keep-it-cold-spray."

This chapter is based on:
Van Hooijdonk, M., Mainhard, M. T., Van Tartwijk, J., & Kroesbergen, E. H. (Under Review). Creative problem solving in primary school students. Submitted for Publication.

Acknowledgement of author contributions:
JvT acquired funding for the study; MvH and TM designed the study; MvH collected the data; MvH and TM planned the data analyses; MvH analysed the data; MvH drafted the manuscript; all authors contributed to critical revision of the manuscript; TM, EK and JvT supervised the study.

Supplementary material can be found at: https://doi.org/10.34894/LALWP0

Abstract
Schools need to foster creative problem solving (CPS) in students from an early age onwards to meet the demand for creative problem solvers in modern society. Insight is needed in the nature of the CPS process and product in primary school students. Findings from a think-aloud study with fourth graders (N = 13) indicated that students' behaviors in response to a CPS task largely matched the CPS model. Findings from a large quantitative study with 4th to 6th grade students (N = 594) showed that the relations of the CPS indicators corresponded with relations found in different age groups. The CPS model can as such be applied in the primary school context and CPS capability is already present in primary school students.

4.1 Introduction

Modern society changes rapidly, and consequently children today grow up with many possibilities and challenges (Craft, 2011). Creative abilities are required in this society to produce both original and useable 'products', within a given social context (Plucker et al., 2004). These products may be artistic expressions such as a painting or a poem, but may also be solutions a primary school student comes up with in class after being confronted with an ill-defined problem. These solutions of students are seen as an expression of everyday (also called "little-c") creativity (Craft, 2011), and are regarded as the outcome of a creative problem solving process (CPS; Isaksen et al., 2011). Creative problem solving (CPS) can, as such, be defined as a real-life creative strategy, in which creativity and general knowledge are combined to solve problems in original but feasible ways (Isaksen & Treffinger, 2004; Long, 2014).

Doing CPS has been shown to benefit both subject matter learning and divergent thinking (Kim et al., 2019; Poddiakov, 2011). In addition, engaging in CPS is associated with primary school students' attitudes towards creativity, active learning, and exploration (Kashani-Vahid et al., 2017; Saxon et al., 2003). Yet, most earlier studies focused on the effects of being involved in CPS on various outcomes, but little is known about the nature of CPS processes in primary school students. CPS is characterized by a process, that consists of several elements that describe distinct CPS behaviors, and a product: a set of potentially creative ideas to solve a problem (Isaksen et al., 2011). These ideas can be scored on multiple indicators to determine the actual creativity of the ideas. The goal of this paper is twofold. First we aim to gain insight into the CPS process in primary school students by studying to what extent the theoretically assumed CPS model can appropriately describe actual CPS behaviors of this young age group in response to a CPS task. Second, we aim to gain insight into the CPS products produced by primary school students, by studying how the four CPS indicators used to determine the creativity of the ideas students come up with relate to each other and to other achievement and creativity tests administrated at school. More insight into the nature of CPS in primary school students could help teachers to foster CPS abilities of students from an early age on and, in this way, could ultimately support students in navigating future uncertainties and possibilities.

The CPS process

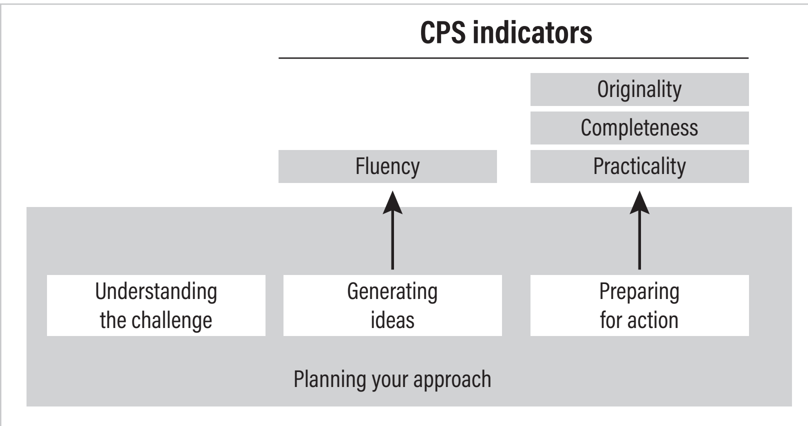
Multiple scholars created frameworks to describe the CPS process and to aid its application (e.g., Altshuller, 1996; Finke et al., 1992; Mumford et al., 1991). Treffinger, Isaksen, and colleagues (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) adapted CPS to the educational context. Their CPS model includes four main elements: understanding the challenge, generating ideas, preparing for action and planning your approach (Figure 1). 'Understanding the challenge' includes three activities: (1) the construction of opportunities by generating broad, brief, and beneficial statements that help set the principal direction for the problem-solving efforts, (2) exploring data by generating and answering questions that pin-point key information, feelings, observations, impressions and questions about the task and (3) framing problems by seeking a specific or targeted question (problem statement) on which to focus subsequent efforts. 'Generating ideas' includes the ability to diverge and think of many, varied and unusual options for responding to the problem (Isaksen et al., 2011). 'Preparing for action' includes the evaluation of ideas and, in this way, identifying the most creative solution to the problem. This element also includes the development of solutions by analyzing and refining promising options. Furthermore, it includes the identification of potential sources of assistance and resistance in practice and other factors that may influence successful implementation of solutions.

These first three main elements should be understood as flexible and dynamic in their ordering: students could apply these in iterative ways switching back and forth between them while solving a problem (Isaksen et al., 2011; Treffinger et al., 2008). Isaksen and colleagues (2011) therefore added a fourth element to the model that includes all efforts to plan and monitor the CPS process called 'planning your approach'. This element also includes determining whether or not CPS is the right choice for the task at stake.

CPS processes have mainly been studied in adult samples (e.g., Peele, 2006; Pringle & Sowden, 2017) and a more complete picture of the CPS thinking processes of primary school students when they solve problems creatively could help us to foster CPS from an early age on. Therefore, in Study 1, we qualitatively explored whether and how the theoretically assumed CPS elements can be used to also describe CPS processes in primary school children. Because the CPS model of Treffinger, Isaksen et al. (2011) model was specifically designed for education, this model was applied in this study.

The CPS product
When students engage in CPS, the creative ideas to solve the problem (i.e., the product) can be described with four CPS indicators: fluency, originality, completeness, and practicality (Figure 1). The number of ideas a student generates during 'idea-finding' is often viewed as an indicator of fluency, that is, how fluent a student is in generating ideas to solve a problem (Isaksen et al., 2011; Reiter-Palmon et al., 2019). Solutions that a student ends up are usually seen as creative when they are unique (i.e., original), well-thought out (i.e., complete) and transferrable to practice (Byrne et al., 2010; Corazza, 2016; Isaksen et al., 2011; Okuda et al., 1991; Piffer, 2012; Reiter-Palmon et al., 2009).

Figure 1
The four elements of the CPS model as based on Isaksen et al. 2011 and the CPS indicators



In multiple studies, the relations between these indicators were investigated. Reiter-Palmon and colleagues (2009) used multiple measures of both originality and quality (encompassing both completeness and practicality) in their study with adults to rate solutions generated in a CPS task. They found that when fluency increases (i.e., more ideas are generated), the number of ideas rated highly on originality increases as well. Between fluency and measures of quality, Reiter-Palmon and colleagues (2009) found some small negative correlations. In addition, measures of originality correlated more strongly with each other than with quality measures, suggesting that originality and quality can be seen as distinctive indicators. Cropley (2006) argued that originality and practicality might even have a negative relationship, because more original ideas tend to be more difficult to translate into practice. Dumas and Dunbar (2014) found comparable positive relations between fluency and originality in their divergent thinking study with adults, concluding that they can be regarded as distinct but positively correlated constructs. The findings described here are based on adult samples and the question is whether the same relations (i.e., a positive relation between CPS fluency and CPS originality, a small negative relation between CPS fluency and measures of CPS completeness and CPS practicality, and a negative relationship between CPS originality and CPS practicality) can be found in a sample of primary school children. When similar relations are found, interventions designed to target specific CPS indicators in adults might be suitable for primary school students as well. Within Study 2, we therefore examined the correlational CPS model as found in adults in a large sample of primary school students.

CPS, divergent thinking & academic achievement
In other samples, CPS showed to be related to divergent thinking and measures of academic achievement (e.g., Gajda et al., 2017; Kwon et al., 2006). When the aim is to embed and assess CPS in primary education, it is important to gain insight in whether these relations are similar for primary school students. This gives us an indication of whether there is an overlap of thinking processes and whether CPS may be a welcome addition to the primary curriculum.

Divergent thinking tasks are often used to get an indication of students' creative potential (Acar & Runco, 2019; Reiter-Palmon et al., 2019). In divergent thinking tasks, students are asked to, for example, think of alternative uses for an object like a brick or a paperclip. These divergent thinking tasks are usually quantitatively assessed on fluency (i.e., how many ideas a subject comes up with), originality (i.e., statistical rarity of ideas) and elaboration (i.e., how detailed ideas are). A critique on such tasks has been that they focus on idea generation in isolation, rather than also including processes demanding both divergent and convergent thinking, such as exploring knowledge, defining the problem and comparing responses to identify the most creative ones that are explicitly part of a CPS task (Cropley, 2006; see also Chapter 2). CPS interventions showed to enhance divergent thinking processes in primary and secondary school students (Kashani-Vahid et al., 2017; Kwon et al., 2006; Scott et al., 2004), suggesting that divergent thinking and CPS are related. Within this study, we investigated how CPS indicators are related to primary school students' divergent thinking outcomes.

Where some authors argue academic achievement and creativity should be related because both require domain-specific knowledge (e.g., Gajda et al., 2017; Schoevers et al., 2020a), other authors argue current academic achievement tests administrated at school leave little room for creativity (e.g., Solomon, 2009), implying that creativity and academic performance are not or even negatively related. To address this debate, Gajda and colleagues (2017) performed a meta-analysis on 120 studies across diverse age groups. They found significant but relatively modest correlations between academic achievement and creativity tests (e.g., $r = .23 - .30$). The studies included in the meta-analysis used a variety of creativity tests such as divergent thinking, drawing, and self-descriptive questionnaires, but the relationship of academic achievement with CPS in primary school students remains unclear. Therefore, we studied how CPS indicators are related to academic achievement as well.

The present studies

The two studies described in this paper aimed to gain insight in the nature of CPS processes of primary school students. This was done by conducting one study on the CPS process (study 1) and one study on the CPS product (study 2).

Study 1

To determine whether the nature of CPS processes in primary school students corresponds with theoretically assumed CPS elements (Isaksen et al., 2011; Isaksen & Treffinger, 2004), we observed primary school students when they were confronted with a CPS task. Because the think-aloud method is especially suitable for studying complex cognitive processes such as creativity and problem solving (Sowden et al., 2020; Van Someren et al., 1994), we used this qualitative approach in Study 1 to answer our first research question:

> *Question 3.1: To what extent do the CPS elements appear when primary school students solve problems creatively?*

To answer this question, we examined to what degree the exhibited behavior matched the proposed processes as presented by Isaksen and colleagues (2011). Because of the explorative nature of this study, no hypotheses were defined.

Study 2

In the quantitative Study 2, 25 4th to 6th grade classes participated and students (N = 594) applied CPS with the same structured task three times for three different problem situations. The ideas of students were rated on the four indicators and a measurement model and structural model were tested in order to answer the second and third research question:

> *Question 3.2: How are the CPS indicators related in primary school students?*
> *Question 3.3: How do these CPS indicators relate to outcomes from a divergent thinking task and to academic achievement?*

Hypotheses

For the relations among the CPS indicators (i.e., the measurement model) we expected CPS fluency and CPS originality to be distinct but positively correlated constructs (Dumas & Dunbar, 2014; Reiter-Palmon et al., 2019). In line with Reiter-Palmon and colleagues (2019), we also expected small negative relationships between CPS fluency and CPS completeness and practicality. Apart from that, we expected CPS originality and CPS practicality to have a negative relationship, because more original ideas may be more difficult to translate to practice (Cropley, 2006).

For the external relations (i.e., the structural model), we expected that CPS fluency would be related to divergent thinking fluency, CPS originality would be related to divergent thinking originality and CPS completeness would we be related to divergent thinking elaboration (Runco & Acar, 2012). We expected CPS originality to be related to divergent thinking fluency as well. In line with the meta-analysis of Gajda and colleagues (2017), we expected academic achievement to have a positive small to moderate association with CPS originality, CPS completeness and CPS practicality. Moreover, we expect the relations between academic achievement and CPS completeness/practicality to be stronger than the relation between academic achievement and originality, because both academic achievement and CPS completeness/practicality tend to rely on convergent thinking processes (De Vink et al., 2022; Webb et al., 2017).

4.2 The Structured CPS task

In both studies structured CPS tasks were used to trigger students creative responses (see Appendix in online materials). We first describe these tasks before proceeding to the methods and results of the qualitative study (study 1) and the quantitative study (study 2).

Problem situation construction

A structured CPS task was constructed and used in study 1 and study 2. In general, creativity is considered to be a partly domain-specific ability (Plucker & Beghetto, 2004; Silvia et al., 2009) and a variety of factors such as creative self-efficacy, prior knowledge and task characteristics tend to influence creativity scores (e.g., Redifer et al., 2021; Reiter-Palmon et al., 2009). Consequently, determining the CPS performance of students with a single task may be difficult, as it relates to only one domain. Therefore, the goal was to construct three different everyday problem situations, related to three different domains, to be solved by the participants. Because the scientific, interpersonal and entrepreneurial domains were found to be distinct creativity domains in several studies (Kaufman, 2012), two problem situations for each of these three domains were selected from Treffinger's practice problems (2000). Based on vignette theory (Poulou, 2001), we modified the six problem situations to fit our research purposes and the age group. As with vignettes, the problem situations described hypothetical scenarios which, although being realistic, did not involve the respondent personally. Six criteria were used to design the problem situation descriptions: (1) The problem situations are open and non-directive, enabling the participant to form his/her own interpretation of the described situation, which is important for creativity; (2) They are concrete and specific about the situation, precisely delineating the situation under investigation; (3) The characters and story described are believable and contain realistic elements; (4) They evoke imagination, feelings and thoughts at the same time; (5) They are written in the third person and contain about the same number of words; (6) They are easily understood and contain short sentences, and the choice of words is suitable for children between 9 and 12 years old.

To select the three problems for our study, five 9-year-old students (4th grade, 3 boys, 2 girls) were asked to rank the six constructed problem situations on complexity, importance, realism, problem-based efficacy (i.e., whether they thought they could solve the problem), and experience. The three problems that were selected ranked lowest on complexity, but highest on realism, efficacy, and experience.

The selected science problem described Lisa and Tom buying ice cream on a hot day, which melted on the way home. The selected social problem referred to a classroom situation, in which Simon gets distracted by his friend Julian all the time. The selected entrepreneurial problem described Aron who desperately wants a new bicycle, but only owns a two-euro coin.

After this step, we asked three experts in educational research to evaluate the three problem situations on the six design criteria. Their feedback was used to enhance the formulation of the problems. In brief, a few sentences were rewritten or shortened and a few words were altered to make the problem situation less specific and easier to understand. The final problem situations included a short story about the problem, and were presented written on paper as well as read aloud to the students by the test administrator.

Steps of the structured CPS task

The structured CPS task included a general introduction of what CPS is and an explanation of the steps of the task. Then, the problem at hand was read aloud to the students. Next, the students were guided through four pre-structured steps of the task. In the first step, students were instructed to explore their knowledge relevant to the problem by listing the knowledge elements they could think of (e.g., the box, the ice, the distance, the temperature for the science problem; cf. Barak, 2013) in a simple mind map. Here, students were told they could list as many elements as they liked. In the second step, students were asked to frame the problem at hand by writing down the problem statement in the form of a question in the middle of the mind map. In the subsequent third step, students received ten minutes to list as many different and original ideas to solve the problem as they could. Students received a sheet of paper with 18 boxes in which to write their ideas, and were told they could list more ideas on the back of the paper if they wished to do so. Because instructions influence creative outcomes (Di Mascio et al., 2018; Nusbaum et al., 2014), the students were explicitly asked to come up with ideas of which nobody else would think. To minimize time pressure, students were not notified about the time left to complete the task. To stimulate the use of knowledge in idea finding, students were asked to look back at their mind map to see what elements they could use in the construction of more ideas. In the final step of the task, top scoring was applied (Silvia et al., 2008). Rather than rating each of a student's ideas (and consequently rating a different number of ideas for each student), the method of top scoring has been developed to assure a fixed number of ratings for every student and a fluency-independent measure (Benedek et al., 2013; Silvia et al., 2008; see also Chapter 2). Students themselves therefore selected their three most creative ideas, which were then considered for scoring. The administrator stated: "A creative idea is an original idea nobody has thought of before, it is complete and solves the problem and can be easily applied in practice." The students were asked to label their self-perceived top three of ideas by numbering them on their sheet of paper, for which purpose there were small boxes beside each idea box.

4.3 Study 1

The structured task was set once for the science problem in a small-scale think aloud study, to study to what extent theoretically assumed CPS elements appear when primary school students solve problems creatively.

Method

Participants

The students that were invited to participate in the study were independently selected by two teachers of two 4th grade classes (N = 42) of one sub-urban school in The Netherlands. Although the CPS task was developed for 4th to 6th grade students, we selected 4th graders for the think aloud study: we reasoned that if it proved possible to elicit CPS processes with 4th grade students, then it was likely 5th and 6th grade students would show these processes as well. Teachers were asked to select students with well-developed verbal communication skills. Although this might have implied these students were slightly more intelligent or creative compared to the rest of their class (i.e., potentially introducing a systematic bias; Batey & Furnham, 2006; Sternberg, 2000), the ability to verbalize thinking processes is regarded as a prerequisite for participating in a think aloud study (Van Someren et al., 1994). Seventeen students were selected by both teachers (9 boys, 8 girls) and were consequently asked to participate in this study. The parents and the students of this group were informed about the study and asked for their consent with a written consent form. Parents received two weeks to decide and were reminded once of the study after the first week. In the end we received consent for 13 students (7 boys, 6 girls). These students were included in the study. Parents of 4 students did not respond and these students were therefore excluded from the study.

Procedure

In order to study whether the students engaged in CPS processes, we asked the students to complete one CPS task while thinking aloud. The think aloud method is the most common approach in studies using problem-solving tasks because problem solving combines two types of reasoning that are relatively easy to verbalize: constructing solutions and constructing justifications for these solutions (Van Someren et al., 1994). Think aloud is regarded as a form of concurrent verbal reporting (Ericsson & Simon, 1980): within a think aloud study, students are individually assessed and instructed to verbalize their thoughts while performing a task. During the administration, it is important that the participants are instructed to report verbal content, but are not asked to explain their thinking (Ericsson & Simon, 1980). If these conditions are satisfied, it is argued that researchers are able to identify the spontaneous use of processes. Although thinking aloud may slightly slow down those processes, the concurrent verbalization of one's thoughts will most likely not interfere with the ongoing thinking processes (Bannert & Mengelkamp, 2008; Ericsson & Simon, 1980), making think aloud a suitable method to study CPS processes. Because original solutions tend to be scarce by nature, we chose to use the science problem situation, as it elicited the most original solutions in an earlier pilot study (see Chapter 2). Furthermore, this problem was easily understood by students. Three trained assistants administered the task. The students were seated outside the classroom in a quiet area. The whole procedure took about 30 minutes for every student. Before the CPS task, the students were shortly trained in thinking-aloud with a simple mathematical word problem (Van Someren et al., 1994; Marc saves money for a new drum set with new drumsticks. He has already saved 559 euros. The drum set costs 1450 euros and the drum sticks 29 euros. How many euros does he still have to save?). The researcher explained that thinking aloud includes saying everything you think while doing a task. The researcher gave an example, e.g.: "I think I need to write down first how much he needs in total, and how I need to compute that.". During the mathematical word problem, the researcher explained what he/she wanted to hear (e.g., Keep on talking, Please tell me what you think?, Please describe what you think now). The word problem task lasted about 5 minutes. Then, the researcher asked whether the concept of thinking aloud was clear before proceeding to the CPS task.

The role of the administrator was a restrained one: interference only occurred when the student stopped talking. The students were prompted to continue thinking aloud when they were silent for five seconds or more by using one of four prompts (Please tell me what you think; keep on talking; what are you thinking?; what are you thinking about?) or by repeating the last sentence or last few words the student expressed. The students were not asked for clarifications or elaborations to educe interference with the cognitive processes involved in creative problem solving (Van Someren et al., 1994).

Table 1

Coding Scheme to Code the CPS Processes of the Students.

Main CPS process	Category	Operational definition	Description of utterances
1. Understanding the challenge *Systematic effort to define, construct, or focus your problem-solving efforts.*	1a. Constructing opportunities	Generating broad, brief, and beneficial statements that help set the principal direction for problem-solving efforts.	Wish: Wouldn't it be nice if… Obstacle: Wouldn't it be awful if…
	1b. Exploring data	Generating and answering questions that bring out key information, feelings, observations, impressions and questions about the task.	Information: knowledge, facts, intelligence, memory, comprehension, recollection. Feeling: emotion, sentiment, awareness, affection, desire, sensitivity, empathy. Observation: Notice, perception, comment, take into action, watch. Impression: Intuitive guess, hunch, image, expectation, belief, notion. Question: inquiry, doubt, perplexity, difficulty, uncertainty, curiosities.
	1c. Framing problems	Seeking a specific or targeted question (problem statement) on which to focus subsequent efforts.	Problem statements: All efforts to state a specific gap between the opportunity you want to create on the future and the current situation.

2. Generating ideas *Generating phase, commitment of extended effort to seek creative possibilities.*	2a. Generating ideas	Coming up with many, varied, or unusual options for responding to the problem.	Idea: A beginning concept, preliminary thought, option or possibility to reach the desired outcome.
	2b. Idea continued		A continuation of the idea in a second utterance.
3. Preparing for action *Make decisions about, develop, or strengthen promising alternatives, and plan for their successful implementation.*	3a. Developing solutions	Analyzing, refining, and developing promising solutions.	Analysis: Screening and analyzing options Decision: Making decisions about alternatives Strengthening: Strengthening tentative solutions
	3b. Building Acceptance	Searching for potential sources of assistance and resistance and identifying possible factors that may influence successful implementation of solutions.	Acceptance: identification of assisters to make success possible. Identification of sources of resistance
4. Planning your approach *Managing the CPS process by thinking metacognitively and applying diagnostic efforts.*	4a. Appraising tasks	Determining whether or not CPS is appropriate for a given task, and whether modifications of one's approach may be necessary.	Appraisal: determining the suitability, worth, and potential effectiveness of applying CPS.
	4b. Designing process	Identify the entry point in the framework, their pathway through the framework, and an appropriate exit point from the framework.	Process: reflections on series of actions or operations to meet the needs of the task. Proposed configurations of steps.

Analysis

The 13 think-aloud protocols were audio-recorded, transcribed, anonymized, and then segmented into utterances. One utterance included one meaningful expression about a (sub)topic. When the student changed topics or when a new element was added, a new segment started. We applied the directed content analysis technique because this method is used to validate or extend a theoretical framework or theory (Hsieh & Shannon, 2005). In our case, we aimed to study whether and to what extent the primary school students' verbalized thinking showed evidence of the CPS elements as described by Isaksen and colleagues (2011) in the CPS model. A coding scheme was therefore prepared, which was used to code the CPS processes of the students (table 1). For the four main elements (understanding the challenge, generating ideas, preparing for action, and planning your approach) of the model of Isaksen and colleagues (2011), we defined categories describing the behaviors to be observed. Next, operational definitions for each coding category were defined using the CPS theory. In a last step, specific descriptions of the utterances belonging to each coding category were defined in line with Isaksen and colleagues (2011). Segments of 8 protocols were coded with this coding scheme in a first phase. Segments that were relevant but that could not be categorized within the initial coding scheme were given a new code. Next, the remaining 5 protocols were coded with this renewed coding scheme. If any relevant segments without a code remained, they were shortly discussed. The program NVivo version 12 Pro was used to code all protocols.

Across the 8 protocols that were coded first, 22 segments (6.4 %) could not be coded with the initial coding scheme. 17 of these 22 segments could be regarded as task-related processes, such as remarks or questions about the task. For these segments, the category 'task-related' was defined for the 'planning your approach' process. From the remaining 5 segments without a code, 2 segments included students asking questions about spelling, 2 segments included a personal story, and 1 segment included an external remark about children playing outside. These segments were defined as irrelevant segments and were as such excluded from further analysis. The renewed coding scheme was used to code the remaining 5 protocols. All segments could be coded with this scheme. The 13 protocols contained 22 to 85 segments (M = 43). A second coder coded 4 randomly selected protocols (student 3, 7, 8, and 11; 156 segments in total). A Cohen's Kappa of 0.85 indicated our raters agreed in the assignment of codes (Landis & Koch, 1977).

Results

Findings will be presented and discussed for every main CPS element as well as for the separate categories (Table 2). With regard to the first element 'understanding the challenge', all students showed utterances that belong to the 'exploring data' category. This means that all students expressed information, feelings, observations, impressions or questions about the problem situation. Examples included "I also know that they are not completely melted, they are mostly soft" (student 3), "And that they feel sad as well" (student 1), "Maybe it's also because of the ice cream itself. How they are packaged. So… But also… How they froze up, how they are made." (student 5). The amount of information explored seemed to differ across students, the number of utterances ranged from 4 to 18 (M = 9.07). For the 'framing problems' category, 12 out of 13 students showed 1 to 3 utterances. Only one student showed 7 utterances that were coded as belonging to the framing problems category. Utterances from this student for instance included: "That their ice pops are melting. So why are the ice pops melting is the question?"; "How did the ice pops melt that quickly?"; "And what could help them?". In other words, all students attempted to seek a problem statement on which to focus subsequent efforts. Although the students were not specifically instructed to 'construct opportunities', 7 students did show utterances that belonged to this category. The brief statements belonging to this category included a wish or an obstacle that shaped the principal direction for problem-solving efforts, for instance: "Yeah right. So the ice pops should not melt. So they have to stay cold. An ice pop that does not melt." (student 2).

With regard to second element 'idea generation', all students showed utterances including ideas to solve the problem. Examples for instance included: "Maybe you can make a bicycle. Then you have a basket in front and then you can see, for example if you are going to eat somewhere in between, you can. If you want it to be heated, you can heat it and if it must remain cold, you can program it to be cold." (student 9), "We could also invent a rain or snow cloud to put above an ice cream box or ice pop. And then it can simply be delivered. So then you get on your bike and then a cloud comes behind you from above the ice cream shop. So when you go to buy it at the store, they ask if you want a rain cloud or a snow cloud. And then you can take it with you. Above your bike or above your bag and then it stays cold." (student 2). The number of ideas generated seemed to vary widely: the students generated a minimum of 2 ideas and a maximum of 24 ideas (M = 10.15). 8 out of 13 students continued their idea in later utterances. These ideas were for instance interrupted by questions about spelling (e.g., "They can call a fantasy animal with magica… How do you write that?" student 7), utterances including knowledge or previous experiences (e.g., "You can light a fire with it. With such a thing and then you think wow! If you put it in the muffler of the car, no more CO_2 comes out of the muffler. And then there is an explosion. I know because I once put a pebble in the muffler and then someone found out and they said wow, this can light a fire. So ice in the muffler and then fire.", student 10) or by the researcher that stimulated the student to continue talking.

With regard to third element 'preparing for action', all students showed some form of evaluation of their ideas to identify their most creative ones. Three students even included some form of refinement (e.g., "It may be, even if it takes a long time, but it can be made. And I don't think other people will come up with a machine that you can put on your bike. And it is practical. And it says, a machine that keeps ice pops cool. And I will add to that, to put it on your bike so that that is clear as well." (student 1). The number of utterances describing this developing solutions category ranged from 1 to 14 utterances (M = 6.31). Although students were not asked to 'build acceptance' for their solutions, four students searched for potential sources of assistance and resistance and identified possible factors that may influence successful implementation of solutions. For example, student 8 said: "Or you can melt chocolate and you add it to it. No, I don't think that will taste good" and student 11 said "Ice pops that never melt would be nice because with hot days at the beach, if it melts and your entire hand is covered, that's not really fun".

For the fourth element 'planning your approach', no students showed utterances belonging to the 'appraising tasks' category, meaning no students seemed to question whether or not CPS was a suitable choice for the CPS task. However, all students did show utterances that could be categorized as belonging to the 'designing process' category, indicating that metacognitive actors are an integral part of the CPS process. These utterances for instance included questions about revising information (e.g., "Can I read the story again?"; student 1), statements that they knew more ideas or were finished thinking about ideas, (e.g., "Oh yes, I know something else. An ice cream that, designing an ice cream that never melts. Uh. Now I don't know anything anymore."; student

12) or utterances including they were ready to move to the next step (e.g., "Now I just have to put the numbers down for the ideas"; student 12). The number of utterances belonging to this category ranged from 3 to 10 utterances (M = 5.31). Aside from this, 11 out of 13 students showed utterances that were categorized as task-related processes. Here, the number of task-related utterances seemed to vary greatly as well. The number of utterances categorized as task-related processes ranged from 1 to 24 utterances (M = 4.55). Utterances for instance included task-specific questions (e.g., "Do I have to write that here?", student 10) and clarification requests.

Although some differences occurred within the main CPS elements across the different categories, the main finding of this think aloud study was that every student showed CPS processes belonging to each of the four main elements of the CPS model (understanding the challenge, generating ideas, preparing for action, planning your approach; Isaksen et al., 2011).

Table 2
Number of Utterances per Category per Student and in Total.

Main CPS process	Category	Student Number													
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**Total**
Understanding the challenge	Constructing opportunities	1	12	-	-	-	4	-	1	-	1	1	-	1	21
	Exploring data	12	5	8	4	9	15	9	4	9	18	5	4	16	118
	Framing problems	7	2	1	2	1	3	2	2	2	3	3	2	3	33
Generating ideas	Generating ideas	10	24	6	8	2	12	8	9	8	18	9	8	10	132
	Idea continued	-	1	-	1	-	19	9	3	2	6	-	2	-	43
Preparing for action	Developing solutions	14	6	6	6	5	14	8	3	6	4	3	6	1	82
	Building Acceptance	-	-	-	-	-	-	-	1	-	6	1	3	-	11
Planning your approach	Appraising tasks	-	-	-	-	-	-	-	-	-	-	-	-	-	0
	Designing process	8	3	4	3	5	6	10	3	3	6	6	5	7	69
	Task related	3	3	1	1	2	2	7	1	1	23	1	5	3	50
Total number of segments		55	56	26	24	22	75	53	27	31	85	29	35	41	559

4.4 Study 2

To study the CPS product in primary school students, the CPS task was administrated three times in a second, large-scale, quantitative study. Here, the aim was to study how the CPS indicators (i.e., fluency, originality, completeness, and practicality) relate to each other and to academic achievement and divergent thinking outcomes.

Method

Participants

Because of the complexity of the model and the hierarchical structure of the data (students nested in classes), an a-priori sample size of 575 students (30 predictors*15 observations; Stevens, 2012) and 25 classes was determined. Therefore, twenty-five classes of 4th and 5th grade students (n = 629; mean age = 10.67) of 14 Dutch primary schools participated in our study. Teachers were asked to participate directly or via their schoolboard and were selected based on their willingness to participate. We deliberately asked teachers of a variety of schools to participate, resulting in a sample of 5 urban schools, 4 sub-urban school, and 4 more rural schools.

Procedure

This second study was part of the larger 'Creative Problem Solvers' data collection effort (see also Chapter 5). This paper describes the first analyses performed on these data. The institutional ethical review board approved the study procedures in March 2018 (nr.: FETC16-058). Teachers were informed about the studies' goals and were asked to read and sign a consent form including three statements. Because the study was non-intrusive, passive informed consent was applied for parents. All parents were however informed with a written message including our contact details. This message included notice that teachers and students would be asked for consent, that parents and students could withdraw from the study at any point. Every school also received information about the study for a newsletter. The students were asked whether they were willing to participate before the start of the data collection. Two parents requested their child not be included in the study and 13 students opted out. The data collection took place in April, May and June 2018, over two sessions. All data were collected at school in the regular classrooms, during school days. Data were collected by the first author and a trained research assistant using an administration protocol. In session one, students completed two CPS tasks (with problem situations from the science and social domain, provided to the students in a random order). In session two, students completed the third CPS task (from the entrepreneurial domain) and the divergent thinking task.

Measures

Creative Problem Solving.

For the CPS originality, completeness, and practicality indicators, two raters rated the top-three ideas as selected by the students on each of the three CPS tasks. The originality, completeness, and practicality of the ideas were evaluated using a modified version of the Consensual Assessment Technique (CAT; Amabile, 1996) on a 5-point scale (0-4). The post-graduate raters received 16 hours of training to understand the CPS tasks, the CPS concepts and the rating schemes involved. Pilot data of 70 students not included in this study were rated and discussed to establish sufficient inter-rater agreement.

Originality.
If the idea was very predictable and commonly known, it received 0 points (e.g., use a cool bag). If the idea clearly reflected an imaginative approach and was completely new, it received 4 points (e.g., make an umbrella that protects you from the sun and has little fans built-in that blow cold air).

Completeness.
If the problem at hand was ignored or just repeated in the idea, it received 0 points (e.g., don't eat ice-cream, make sure it does not melt). If all the steps of the idea were explicitly described, it received 4 points (e.g., put a fridge on your bike that is powered by a dynamo and put the ice in when you leave the supermarket).

Practicality.
If the idea was impossible in practice, it received 0 points (e.g., go to a wizard to refreeze the ice-cream with magic). If the idea could be implemented in practice right away, it received 4 points (e.g., buy new ice-cream and eat the ice-cream immediately just outside the store on a bench).

The ratings for originality, completeness, and practicality were averaged across the two raters. Intraclass correlation coefficients (consistency; single measures; Cicchetti, 1994) were calculated for the top 1 to top 3 scores for all three indicators to check for inter-rater agreement. The ICCs for originality ranged from .75 to .87, indicating excellent agreement. The ICCs for completeness ranged from .68 to .80, indicating good to excellent agreement. The ICCs for practicality ranged from .79 to .89 again indicating excellent agreement.

Fluency.
To assess CPS fluency, two raters counted the total number of different ideas listed. Ideas that could not be interpreted or were listed twice, were excluded. Because the raters disagreed in less than 1% of the cases, these cases were discussed to gain a single fluency measure for every student.

Divergent thinking.
One divergent thinking task from the Runco Creativity Assessment Battery (rCAB; Runco, 2011) was applied in this study. This assessment battery is comparable to other divergent thinking assessments such as the Torrance Test of Creative Thinking (TTCT; Torrance, 1972). The task included a picture of a toothbrush and a page full of lines for writing. Students were asked to list as many different uses for a toothbrush as they could. Again, students were explicitly asked to come up with ideas of which nobody else would think. The task lasted 8 minutes and students were not notified about the time they had for the task.

Fluency.
To obtain scores for divergent thinking fluency, the total number of different ideas listed was used. If ideas were listed twice, the second idea was excluded.

Originality.
To obtain scores for divergent thinking originality, a lexicon was created. In this lexicon, the common ideas were clustered to determine which ideas deserve points for being more unique. Ideas were clustered when they had the same meaning, they meant the same but were written

as an action and an object respectively, objects could be replaced with an object from the same category that could be regarded as unique for the student's situation or when they had the same meaning but differ in the amount of detail. A second researcher independently checked whether ideas were correctly clustered. After this check, 18 out of 570 categories (3.15%) were included in a different cluster. Clusters mentioned by fewer than 1% of the students received 1 originality point. To obtain a fluency-independent measure, the originality score was divided by the number of ideas (Reiter-Palmon et al., 2019).

Elaboration.

To obtain scores for divergent thinking elaboration, the amount of detail presented for the ideas, we counted the total number of words a student used for all the ideas and divided it by the total number of ideas (fluency score).

Academic achievement.

Scores from two standard Dutch achievement test batteries (Cito and Boom; Janssen et al., 2010; Tomesen et al., 2019; Van Vugt et al., 2019) were used as a measure for academic achievement. Tests for mathematical ability and reading comprehension were selected, because they give a broad image of a student's academic achievement and were administrated at every school in our sample. We calculated learning efficiency percentage scores for the tests for mathematical ability and reading comprehension by dividing the student's test score by the national mean test scores for every grade (as provided by the test developer) and multiplying this by 100. This means that a student that scored exactly average for his/her age received a learning efficiency percentage of 100%.

Analysis

CPS measurement model.

To study how the CPS indicators (i.e., fluency, originality, completeness, & practicality) relate to each other, we conducted a Confirmatory Factor Analysis (CFA) to test our CPS measurement model with Mplus 8.5 (Muthén & Muthén, 2012). The separate top-three scores for originality, completeness and practicality were first loaded on the three measures of CPS within the tasks. The separate top 1, 2, and 3 scores for originality, completeness, and practicality came from one and the same idea. As a consequence, these scores were expected to show shared method variance (Cole et al., 2007). As such, residual errors belonging to the scores of the same selected ideas were allowed to covary. Next, the first-order latent variables for originality, completeness, and practicality were loaded on the second-order factors of originality, completeness, and practicality across tasks (Figure 2). The students' fluency scores on the three tasks were also loaded on a latent fluency factor across problem domains. Correlations of the four main latent factors (fluency, originality, completeness, and practicality) were included in the specification of the model to determine whether the CPS indicators relate to each other as expected.

As is often the case in creativity studies, highly original ideas were relatively scarce and practical ideas were relatively common within this study. As a result, the distribution of these variables was non-normal. To take this non-normality and the clustering of students in classes into account, type = complex was applied in Mplus as the method of analysis. The Robust Maximum Likelihood (MLR) estimator used with this method is robust to non-normality and non-independence (Byrne, 2013).

The measurement model was assessed based on the following goodness-of-fit indices: the comparative fit index (CFI), the Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA). the standardized root mean square residual (SRMR), and the chi square/df ratio (Hu & Bentler, 1999; Marsh et al., 2004).

Structural model.
To determine how the CPS indicators related to divergent thinking and academic achievement scores from primary school students, we tested a structural model. Within this model, the scores for mathematics and reading comprehension were first loaded on the latent academic achievement factor (Figure 3). Next, the CPS fluency score was correlated with the divergent thinking fluency score. The CPS originality score was correlated with the divergent thinking fluency score, the divergent thinking originality score, and the latent academic achievement factor. The CPS completeness score was correlated with the divergent thinking elaboration score and the latent academic achievement factor as well. CPS practicality was only correlated with academic achievement. Because the divergent thinking measures were taken from the same divergent thinking task, these variables were allowed to covary. Again, we assessed model fit with the CFI, TLI, RMSEA, SRMR and chi square/df ratio indices. Standardized regression coefficients and explained variances (R^2) of the CPS variables were reviewed as well.

Results

Missing data.
Though 627 students and their parents were selected to participate in the study, partly to fully available data of 594 students could ultimately be used in the analysis. Thirty students did not give consent, asked to stop the administration during the first assignment or were a absent or ill during both days of the administration. These students were consequently excluded from the study. Data of 6 students were excluded due to behavioral problems that interfered with the assignment or due to unscorable tasks. A small group of students (1.59 – 2.75% per CPS task) did not define a top-3, or their top-3 could not be interpreted. Pilot study results indicated that students are quite able to define a top 3 scoring highly on the CPS aspects, as scored with our rating scheme (see Chapter 2). Therefore, we defined a top-3 for these tasks ourselves. The top 3 scores of these tasks were rated by the same two raters. For students that were only able to think of 1 or 2 ideas, only scores for these ideas were included. This resulted in a slightly different sample size for every measure (Table 3; Table 4). Academic achievement data was missing for mathematics for 15 students (2.51%) and for reading comprehension for 21 students (3.53%). No outliers were detected. The MLR estimator used for the analyses enabled us to estimate the model using all the available data.

Table 3
Mean Top Scores for CPS Originality, Completeness, and Practicality across Raters for the Science, Social, and Entrepreneurial Tasks.

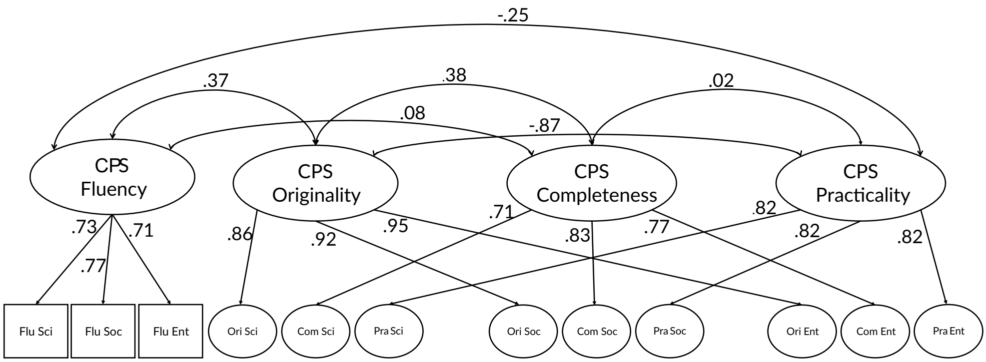
		CPS Originality				CPS Completeness				CPS Practicality			
	N	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
Science task													
Top 1	566	2.07	1.05	1.00	5.00	3.24	0.81	1.00	5.00	4.48	0.86	1.00	5.00
Top 2	553	1.97	0.99	1.00	5.00	3.14	0.69	1.00	5.00	4.60	0.78	1.00	5.00
Top 3	499	1.94	0.91	1.00	5.00	3.05	0.73	1.00	5.00	4.60	0.78	1.00	5.00
Social task													
Top 1	565	2.02	0.95	1.00	5.00	3.29	0.68	1.00	5.00	4.27	0.85	1.00	5.00
Top 2	540	1.93	0.89	1.00	5.00	3.23	0.62	1.00	5.00	4.30	0.82	1.00	5.00
Top 3	496	1.93	0.84	1.00	5.00	3.15	0.61	1.00	4.50	4.33	0.76	1.00	5.00
Entrepreneurial task													
Top 1	546	2.02	0.95	1.00	5.00	3.21	0.74	1.00	5.00	3.78	0.93	1.00	5.00
Top 2	538	1.93	0.89	1.00	5.00	3.11	0.68	1.00	5.00	3.82	0.86	1.00	5.00
Top 3	515	1.93	0.84	1.00	5.00	3.04	0.69	1.00	4.50	3.89	0.83	1.00	5.00

Table 4
CPS Fluency Scores for the Science, Social, and Entrepreneurial Tasks, Divergent Thinking Scores, & Academic Achievement Scores

	N	M	SD	Min	Max
CPS Fluency					
Science	572	5.73	2.97	1.00	17.00
Social	568	5.75	3.31	1.00	16.00
Entrepreneurial	550	6.76	3.51	1.00	17.00
Divergent Thinking					
Fluency	558	11.27	6.85	0.00	36.00
Originality	547	0.10	0.12	0.00	1.00
Elaboration	547	2.75	1.31	1.00	12.00
Academic Achievement					
Mathematics	585	98.51	11.70	51.09	135.63
Reading Comprehension	579	97.59	25.49	12.50	209.38

The CPS measurement model.
The CFA of the CPS measurement model with nine first-order and four second-order latent variables showed an adequate to good fit between the model and the observed data (CFI = 0.95; TLI = 0.94; RSMEA = 0.03, 90% CI [0.03-0.04]; SRMR = 0.05; chi square/df ratio = 568.24/363 = 1.57). No post-hoc modifications were conducted. Standardized and unstandardized parameter estimates for within-task first-order latent originality, completeness, and practicality variables and residual correlations are provided in the appendix (see online materials). The residual errors and correlations were checked. No negative residual errors were detected. The patterns of the residual correlations corresponded within the tasks, illustrating the presence of shared method variance (see the online materials). Standardized factor loadings and correlations of the latent fluency factor and the second order (across-task) latent variables are provided in Figure 2.

Figure 2
Results CPS measurement model



Note. Flu = CPS Fluency; Ori = CPS originality; Com = CPS completeness; Pra = CPS Practicality; Sci = Science problem situation; Soc = Social problem situation; Ent = Entrepreneurial problem situation.

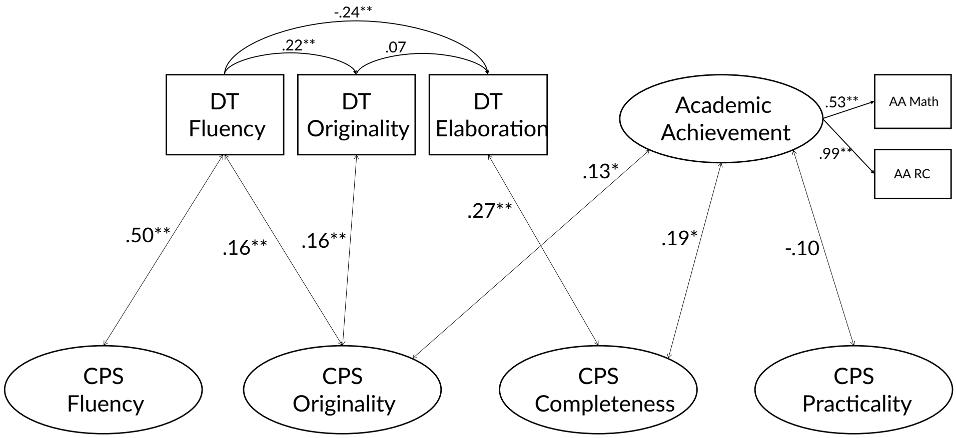
All within-task CPS factors loaded significantly on the across task second order latent CPS factors ($p < .001$ for all factor loadings). The CPS fluency measures of the three tasks loaded significantly on the CPS Fluency factor as well ($p < .001$). Overall, correlations between the factors were mostly in line with what we expected based on the literature. Positive moderate correlations were found between CPS Fluency and CPS originality and CPS Originality and CPS Completeness ($r = .37$, $p < .001$ and $r = .38$, $p < .001$ respectively). A large negative correlation was found between CPS Originality and CPS Practicality ($r = -.87$, $p < .001$). A small negative correlation was found between fluency and practicality ($r = -.25$. $p = .001$). No significant correlations between fluency and completeness ($r = .08$, $p = .25$) and completeness and practicality ($r = .02$, $p = .86$) were found.

Structural model.
The analysis of the structural model also showed an adequate to good fit between the model and the observed data (CFI = 0.93; TLI = 0.92; RSMEA = 0.03, 90% CI [0.03-0.04]; SRMR = 0.06; chi square/df ratio = 817.82/510 = 1.60). Again, no post-hoc modifications were conducted. Standardized regression coefficients and factor loadings for the academic achievement factor ($p < .001$ for both academic achievement factor loadings) are provided in Figure 3.

4

Divergent thinking fluency was significantly related to CPS fluency ($r = .50$, $p = < .001$). This correlation can be classified as large (Cohen, 1988). CPS originality was significantly related to divergent thinking fluency, divergent thinking originality and academic achievement ($r = .16$, $p < .001$; $r = .16$, $p < .001$; $r = .13$, $p = .032$ respectively). These correlations can be classified as small (Cohen, 1988). CPS Completeness was significantly related to divergent thinking elaboration and academic achievement ($r = .27$, $p < .001$; $r = .19$, $p = .002$). These correlations can be classefied as small as well (Cohen, 1988). No significant relation was found between CPS practicality and academic achievement ($r = -.10$, $p = .26$).

Figure 3
Results of the structural model for the relations between the CPS indicators, the divergent thinking (DT) outcomes and the latent academic achievement (AA) factor.



*Note. *p < 0.05, **p < .001; Math = Mathematics; RC = Reading Comprehension*

4.5 Discussion

Schools need to foster CPS skills in students from primary education onwards to meet the growing demand for creative problem solvers in modern society. In order to do so, more insight is needed in the CPS processes of this young age group. The two studies on the CPS process (study 1) and CPS product (study 2) we conducted aimed to gain insight into the nature of CPS in primary school students.

The CPS process

The main finding of the think-aloud study of the CPS process (Study 1) was that each student engaged in the CPS processes belonging to the four main elements of the CPS model (understanding the challenge, generating ideas, preparing for action, planning your approach; Isaksen et al., 2011). This indicates that the CPS model and its main elements can be used to describe CPS processes in primary school students. Some differences between students occurred for the distinct categories within the four elements of the CPS model. For instance for the 'exploring data' category of the understanding the challenge element, the amount of information students explored varied greatly. Possibly, students differed in their prior knowledge or differed in how fluent they were in retrieving this data from their memory (Barak, 2013). For the 'framing problems' category, the variation was much smaller: only one student showed more than three expressions within this category. In the CPS task, students were instructed to explore many knowledge elements, but were asked to formulate only one problem statement. Earlier studies with adults on both CPS and divergent thinking task instructions found that instructions influence eventual creative outcomes (e.g., Di Mascio et al., 2018; Nusbaum et al., 2014). In future research, it is therefore important to pay attention to variations of instructions of the CPS task with primary school students and how these instructions may impact CPS thinking processes and eventual CPS outcomes for this age group. Nevertheless, we conclude that the CPS model of Treffinger, Isaksen, and colleagues applied in this study (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) can be used to describe the creative problem solving process of primary school students.

The CPS product

We studied fluency, originality, completeness and practicality as four elements underlying the CPS product - the ideas students come up with – in a large-scale quantitative study (Study 2). The main findings were that the four CPS indicators (fluency, originality, completeness, and practicality) are distinguishable in the ideas of primary school students and that the relations of these CPS indicators largely matched relations found in other age groups.

For instance, the moderate positive correlation between CPS fluency and CPS originality that was found in our sample of young students corresponds with the relations other authors found with other creativity tasks in older age groups (Reiter-Palmon et al., 2019). This suggests similar structural relations of the four CPS elements for primary school students. In addition, this indicates that CPS originality and CPS fluency, like divergent thinking originality and divergent thinking fluency (Dumas & Dunbar, 2014), can be regarded as distinct but related constructs.

In previous studies in older age groups, the completeness and practicality indicators were assessed as one single construct to measure overall quality or effectiveness (e.g., Corazza, 2016; Reiter-Palmon et al., 2019). Within the current study, CPS completeness and CPS practicality were assessed separately to obtain distinct measures for the amount of detail and the practical feasibility of the ideas. Interestingly, CPS practicality was negatively related to CPS fluency and CPS originality (as was expected based on Reiter-Palmon et al. (2009), but CPS completeness showed a moderate positive correlation with CPS originality. This indicates that, at least in our sample of primary school students, the CPS completeness and practicality indicators can be regarded as distinct constructs. Consequently, they may be best measured separately in this young age group. An alternative explanation is that primary school students did not take practicality into account when they selected their top-3 of most creative ideas, as was suggested in the study described in Chapter 2. More research is necessary to explore this hypothesis.

Although a strong correlation (r = .50) between CPS Fluency and divergent thinking fluency was found, the correlations of CPS originality and CPS completeness with the corresponding divergent thinking outcomes were rather small. Measurement-specific factors may play a role here (e.g., Barbot et al., 2016a). Our way of conceptualizing and measuring fluency in the CPS context was largely in line with how fluency is commonly assessed in divergent thinking tasks (i.e., counting the number of ideas generated). This probably resulted in more overlap and thus explained variance. For originality and completeness, our individual-response qualitative way of conceptualizing and scoring with the consensual assessment technique differed from this quantitative response-set ways of scoring originality and elaboration of the divergent thinking tasks, probably resulting in less shared variance. To select creative indicators and gain insight in ways of scoring for future studies, we recommend authors to consult the work of Reiter-Palmon and colleagues (2019).

In short, the study on the CPS product showed that the CPS indicators used to describe the creativity of ideas in older age groups can be used to describe the creativity of ideas of primary school students as well.

Limitations

Where most earlier studies focused on the effects of being involved in CPS on various outcomes, this study uniquely focused on the nature of CPS processes in primary school students in both a qualitative and quantitative way. Because the think aloud sample was different from the sample in the large-scale administration, comparisons between the two studies should only be done with caution. Although the applied sample of the think aloud study (Study 1) allowed us to investigate CPS processes in greater detail, more students should be added to be able to conclude whether CPS processes lead to corresponding CPS behavior. Furthermore, the young age of the students in Study 1 might have made it difficult for them to verbalize every thought during the think aloud study. Although we put effort into maximizing the students' ability to achieve this by selecting students with good verbal skills, by practicing thinking aloud with a word problem, and by reminding them to verbalize throughout the task, findings should still be interpreted with care. In addition, the overall number of ideas that students came up with in Study 1 seemed larger (M = 10.15) than in the larger scale Study 2 (M = 5.73). This could be due to the selection of the students, but it could be the case that during the idea generation process, some ideas do not end up on paper. More research is necessary to explore how this process works with primary school students.

In the quantitative study (Study 2), top-scoring was applied. Although this method was validated in multiple studies (Benedek et al., 2013; Silvia et al., 2008) and tested with primary school students (see Chapter 2), it could still be the case that students missed creative ideas in their selection, which as a consequence were not considered for scoring. In addition to that, only the first two CPS tasks in Study 2 were provided to the students in a random order. Although the third CPS task administered during the second occasion stemmed from a different domain, (i.e., the entrepreneurial domain) and as such demanded different knowledge, a learning effect might have occurred. Even though the study did not focus on a comparison of domains, the three CPS tasks should be randomized in the future to limit the impact of this learning effect on the results.

Conclusion and practical implications

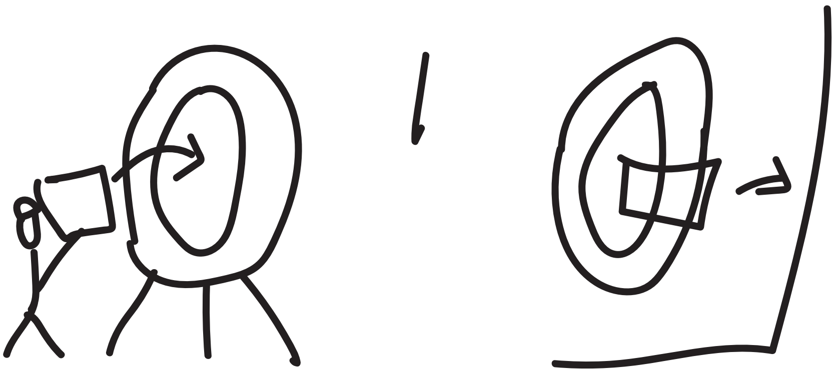
Overall, we conclude that both the CPS model (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) and the CPS indicators can be used to describe the CPS process and product in primary school students. This implicates that the CPS model can be applied in the primary school context and that CPS capability as found in older samples appears to be already present in this young age group.

A structured CPS task was applied to trigger the CPS processes across three problem situations. These kind of CPS tasks can be a welcome addition to the repertoire of creative activities with primary school students, as they are directly connected to everyday life (Piffer, 2012; Zeng et al., 2011) and include processes such as exploring knowledge, defining the problem, and comparing responses to identify the most creative ones, besides the processes of idea generation. Our results indicated that CPS and academic achievement are related, but that they are distinct from each other and therefore seem to demand different cognitive processes. Although more in-depth research is necessary to determine the specific differences and overlap between the two, these results suggest that adding these kind of CPS tasks to the curriculum does not imply unnecessary overlap with other academic activities. In fact, it might even add a low-stakes alternative to high-stakes testing, since CPS tasks require the exploring of diverging solutions and not a single, correct answer (Baer & Garrett, 2017; Runco et al., 2017). By regularly doing these kind of whole-group tasks with primary school students and reviewing their CPS outcomes, it may be easier for teachers to determine and ultimately stimulate students' individual CPS abilities. Consequently, they may be better able to embed CPS and suitable individual CPS interventions in their everyday teaching practices (Beghetto, 2013; Kaufman et al., 2016a). Offering this kind of CPS activities on a regular basis may support primary school students to apply their knowledge in a more flexible and creative way, from an early age onwards.

4

5

CAN ELEMENTARY SCHOOL TEACHERS ASSESS STUDENTS' CREATIVE PROBLEM SOLVING ABILITIES?



"Simply use a teleporter to beam the ice pops home."

This chapter is based on:
Van Hooijdonk, M., Mainhard, M. T., Kroesbergen, E. H., & Van Tartwijk, J. (Under Review). Can elementary school teachers assess students' creative problem solving abilities? Submitted for Publication.

Acknowledgement of author contributions:
JvT acquired funding for the study; MvH, TM and JvT designed the study; MvH collected the data; MvH and TM planned the data analyses; MvH analysed the data; MvH drafted the manuscript; all authors contributed to critical revision of the manuscript, TM, EK and JvT supervised the study.

Supplementary material can be found at: https://doi.org/10.34894/46OR4F

Abstract
In this study, the validity of teacher assessments of three creative problem solving (CPS) abilities, i.e., the ability to produce original, complete, and practical ideas, was studied. 610 students from 4th to 6th grade were assessed with two CPS tasks. The alignment between teachers (N = 26) and trained raters (N = 2) was studied for both absolute assessments and the relative ordering of students' scores. The main finding of this study was that elementary school teachers were able to assess CPS in a relative way. They were also able to assign absolute scores for originality, but not for practicality or completeness. Contrary to former studies, teachers' CPS assessments in the current study were not affected by students' gender or academic achievement.

5.1 Introduction

Children today face numerous uncertainties and possibilities, because of the many smaller and bigger changes in society (Craft, 2011). To enable them to cope with these unknown problems of tomorrow, schools need to foster the creative problem solving abilities of students, preferably from an early age onwards (Isaksen et al., 2011; Kim et al., 2019). Creative problem solving (CPS) is a form of everyday creative ability in which students use both subject knowledge and creative skills to solve everyday problems (Craft, 2011; Isaksen & Treffinger, 2004; Okuda et al., 1991). To be able to foster CPS, teachers need to have insight in students' creative abilities (Bolden et al., 2020). Unfortunately, determining students' CPS abilities is a challenge in general, and probably even more challenging for teachers in specific. Student performance is affected by a variety of factors such as specific task characteristics (e.g., perceived importance or complexity of the problem task; Reiter-Palmon et al., 2009) and task- or domain specific creative abilities (e.g., differences in domain-specific knowledge or skills; Barbot et al., 2016b). As a consequence, student performance can vary between different tasks. On top of this, teachers experience difficulties when they have to assess creative abilities and seem to be biased in their judgements (Benedek et al., 2016; Bolden et al., 2020). For instance, students that functioned well at school tended to be favored in teachers' assessments of creativity and the creativity assessment of boys tended to be more accurate than those of girls (Gralewski & Karwowski, 2013). It is, however, not yet studied how potential assessment biases affect scores on CPS tasks. Analyzing teacher assessments of students' CPS abilities could provide relevant information about how teachers judge CPS in the classroom and might also enhance the measurement of CPS in everyday educational practice. Therefore, the goal of this study was to determine the validity of teachers' CPS assessments to gain insight in whether teacher ratings can be used in the classroom to determine elementary students' CPS abilities.

5

CPS in education

Interest in CPS in elementary education is growing, because it can be easily connected to the problems a student may face in daily life as well as to factual knowledge learned in school (Kim et al., 2019). In addition, engaging in CPS could be beneficial for elementary students' divergent thinking, attitudes towards creativity, active learning and exploration (Kashani-Vahid et al., 2017; Kim et al., 2019; Saxon et al., 2003). One way to nurture CPS in the classroom is by teaching students to tackle everyday challenges by trying to understand the underlying problem, then generating multiple creative ideas to solve the problem and selecting the most original, yet feasible, ideas to put into practice (Isaksen et al., 2011; Sophonhiranrak et al., 2015). Although teachers acknowledge the importance of fostering creative abilities in students (Kasirer & Shnitzer-Meirovich, 2021), they still struggle to implement CPS in their teaching practices (Kim et al., 2019). Davies and colleagues (2014) studied how teachers can be best supported in facilitating creative learning environments. Among other resources, school-based creativity assessments to determine students' needs could improve teaching for creativity practices. Indeed, creativity assessment skills are viewed as an important prerequisite for fostering creative abilities in students (Benedek et al., 2016; Bolden et al., 2020; Cropley, 2001). To improve CPS practices in education, this study therefore focuses on teachers' assessments of students' CPS abilities.

Assessing CPS

Most often, students' CPS abilities are determined by assessing the ideas a student has come up with. These ideas are regarded creative when they are original, well-thought out (i.e., complete) and transferrable to practice (Corazza, 2016; Okuda et al., 1991; Reiter-Palmon et al., 2009). Although these three assessment criteria (i.e., originality, completeness, and practicality) may be quite straight forward as such, it is still difficult to determine students' CPS abilities. A variety of factors can impact the eventual scores of students on creativity tasks, such as motivational factors, rater biases and task characteristics (e.g., the perceived importance and complexity of the task; Amabile & Pillemer, 2012; Conti et al., 1996; Piffer, 2012; Reiter-Palmon et al., 2009). In creativity research, raters are therefore extensively trained to qualitatively assess creative outcomes. For example, Amabile (1996) proposed the Consensual Assessment Technique (CAT): Multiple, thoroughly trained raters (or experts in the field) assess the creativity of a product and when consensus is reached, the level of creativity is determined. This leads to at least two issues with teachers assessing CPS in students: Firstly, there are practical boundaries. Teachers are - in general - not experts in creativity assessment and their time for CPS assessment training is limited (Davies et al., 2014; Troman & Woods, 2001). Moreover, asking multiple teachers to assess the ideas of every single student is simply too time-consuming. Indeed, multiple studies illustrate that teachers struggle to properly assess student outcomes of a variety of creativity tasks (e.g., Beghetto et al., 2011; Gralewski & Karwowski, 2013; Urhahne, 2011). Gralewski and Karwowski (2013) for instance studied whether and to what extent teachers were able to determine students' creative abilities. They asked multiple teachers to rate the creativity of each student and compared these ratings with students' self-reported creative attitude, creative activity and creative performance measured by a trained rater on a creative drawing test. In general, the accuracy of the teachers' ratings of students' creativity was low. The latent factor of students' creativity did predict teachers' ratings, but only moderately. Urhahne (2011) studied whether teachers determined elementary students' creativity correctly, to be able to identify potential gifted students. Standardized testing procedures were used to assess creativity and teachers filled in a rating scale. Teachers failed in predicting students' creativity as measured with the test. Beghetto and colleagues (2011) conducted two studies in science and math, exploring the relationship between elementary students' self-judgments of creative ability and teachers' ratings of students' creativity. In both studies, students' self-judgments of creativity in science only predicted a small proportion of variation in the teachers' ratings of students' creative expressions. These results indicate that teacher assessments of creativity may be partly aligned with students' self-judgements of creativity, but are most-often not aligned with creativity measures scored by trained raters. It should be noted that the studies described here primarily focused on divergent thinking tests that are most often used to determine the creative potential of students. To the best of our knowledge, the alignment between teacher assessments and trained rater assessments of CPS tasks has not yet been studied.

A second issue is that teachers, in contrast to trained raters, are not objective raters because they know their students. Consequently, teacher assessments of CPS may be influenced by students' characteristics apparent in class (Gralewski & Karwowski, 2013; Walton & Kemmelmeier, 2012). This could lower the validity of teachers' assessments of children's creativity. Gralewski and Karwowski (2013) for instance studied to what extent the assessment of creativity was biased by students' performance, measured with students' GPA, behavioral grades and participation in School Olympics. Results indicated that students that functioned well at school were perceived as more creative by their teachers. The authors also found a gender bias: Only for male students the prediction of creativity levels was adequate. Sommer and colleagues (2008) found the opposite: teachers and parents were better able to predict the creativity of girls and showed to be more biased in their creativity assessments towards boys. For CPS specifically, these biases have not yet been studied. Therefore, within this study, biases in teachers' CPS assessment are explored as well.

Absolute and relative assessment

In creativity research, even when raters are thoroughly trained, every creativity task still tends to yield different absolute creativity scores (Long & Pang, 2015; Reiter-Palmon et al., 2009; see also Chapter 3). As a consequence, specific norms can be constructed for every single CPS task. This, however, could make the assessment of CPS by teachers even more difficult and labor intensive. Therefore, within this study, we focused not only on the absolute (i.e., raw) scores of students, but also on student's relative performance (i.e., group mean centered scores) across CPS tasks. When assessment focuses on the relative creative performance of students, a student's creativity score is determined by comparing it with the scores of the other members of a group (Cseh & Jeffries, 2019; Leikin, 2009). In this way, the relative assessment of CPS for instance focuses on whether a student is able to think of more creative solutions over time compared with this student's peers. To study what approach to the assessment of CPS would suit current educational practices best, absolute and relative assessment scores of teachers and trained raters are compared in this study.

The present study

In short, two difficulties complicate the assessment of CPS. On the one hand, teachers tend to struggle with assessing creative abilities in students and are possibly biased in their judgements. On the other hand, it appears to be complicated to develop standardized norms for CPS tasks in general. Still, it is important to explore how CPS can be assessed by teachers, in order to promote creativity in elementary school classrooms.

The goal of this study therefore was to determine whether and in what way teacher assessments could be used to gain a valid estimate of elementary students' CPS abilities.

Specifically, we addressed the following research question:

> *Question 4.1: Can elementary school teachers assess students' creative problem solving abilities?*

In order to answer this question, we evaluated whether and to what degree teacher assessments of three CPS abilities (i.e., the ability to produce original, complete, and practical ideas) are aligned with trained rater assessments. Students' gender and academic achievement scores were studied as well to gauge potential biases in teacher assessments of CPS. Both the absolute CPS scores and the relative ordering of students were compared between teachers and trained raters.

5.2 Method

Participants

Twenty-six classes of 4th to 6th grade students (N = 650; M age = 10.66, 55% female) and their teachers (N = 26, M age = 41.23, 80.8% female; M teaching experience = 15.15 years) of 15 Dutch elementary schools were asked to participate in our study. Teachers were sampled for convenience. To ensure our sample of students would be diverse in terms of cultural background, we deliberately asked teachers of a variety of schools to participate resulting in a sample of 6 urban schools, 4 sub-urban school, and 4 more rural schools. Parents were informed about the study and its consent procedure. Teachers and students were asked for consent and students could withdraw from the study at any point. One parent did not give consent and 29 students did not want to participate in the study when filling in the consent form, were ill/absent on both occasions, or asked to stop the administration during the first assignment and were as such excluded from the study. This means that a total of 620 students eventually participated in the study.

Materials

CPS exercise

A four-step creative problem-solving task for elementary education as deployed in Chapter 4 was used. Within this exercise, different problem situations are presented to students, which they are asked to solve creatively. Because of the task-specific nature of creativity, CPS scores were compared across two CPS tasks, stemming from two domains (Kaufman, 2012; Oral et al., 2007): One problem from the science domain (Figure 1) and one problem from the social domain (Figure 2).

Figure 1

Problem Situation 1, Stemming from the Science Domain

> Today, the weather is very hot. Lisa and Tom feel like eating ice cream. They cycle to the supermarket and buy a box with ice pops. When they get home and open the wrapper, the ice pops appear to be completely soft. The ice drips down the stick. The ice pops have started to melt!

Figure 2

Problem Situation 2, Stemming from the Social Domain

> Simon sits next to his friend Julian in class. Simon is often distracted by Julian, causing him to miss an important part of class. Today, Simon again does not get his school work done because Julian disturbs him. He is fed up! But Simon doesn't want to lose Julian as a friend either. He likes hanging out with him.

The exercise included four steps. In the first step, students were instructed to apply fact finding by listing the knowledge elements they could think of (e.g., the box, the ice, the distance, the temperature for the science problem) in a simple mind map, which is commonly used in Dutch elementary education. Here students were told they could list as many elements as they liked. In the second step, students were instructed to apply problem finding by writing down the problem statement in the form of a question in the middle of the mind map. Subsequently, in the third step, students received ten minutes to list as many different and original ideas as they could, to solve the problem. Students received 18 boxes to note down their ideas and were told that they could list more ideas on the back of the paper if they wished to. Because instructions influence creative outcomes (Nusbaum et al., 2014; Runco et al., 2005), the students were explicitly asked to come up with ideas nobody else would think of. To minimize time pressure, students were not informed about the time left to complete the exercise. To stimulate the use of knowledge in idea finding, students were stimulated to consult their mind map to see what they could use for formulating more ideas. In the fourth and final step of the exercise, top-scoring was applied (Silvia, et al., 2008). Here, students chose their three most creative ideas. Students were asked to take originality, completeness and practicality into account: The administrator stated: "A creative idea is an original idea nobody has thought of before, it is complete and solves the problem and can be easily applied in practice." The students were asked to highlight their top 3, by noting 1, 2 or 3 in a small box behind their selected ideas. In the current investigation only the top-1 ideas were included to keep the teachers' time investment limited.

Measures

Trained rater assessment of CPS

To assess the originality, completeness, and practicality of the ideas, two trained raters (the first author and a graduate student with expertise in creativity research) rated the most creative ideas, as selected by the students. The originality, completeness, and practicality of the ideas were evaluated using a modified version of the CAT (Amabile, 1996). The raters received at least 16 hours of training to understand the CPS exercises, the CPS concepts and the ratings schemes involved. Pilot data of 70 students were rated and discussed to establish sufficient inter-rater agreement.

CPS originality.

Originality was rated with a rating scheme by the two raters on a five-point Likert scale. If the idea was very predictable and commonly known, it received 1 point (e.g. use a cool bag). If the idea clearly reflected an imaginative approach and was completely new, it received 5 points (e.g., make an umbrella that protects you from the sun and has little fans built-in that blow cold air).

CPS completeness.

Completeness was rated on a five-point Likert scale. If the problem at stake was ignored or just repeated in the idea, it received 1 point (e.g., don't eat ice-cream, make sure it does not melt). If multiple steps towards a solution were included in the idea, it received 5 points (e.g, put a fridge on your bike that is powered by a dynamo, put the ice in when you leave the supermarket and cycle home).

CPS practicality.

Practicality was also rated on a 5-point Likert scale. If the idea was impossible in practice, it received 1 point (e.g., go to a wizard to refreeze the ice-cream with magic). If the idea could be transferred to practice right away, it received 5 points (e.g., buy new ice-cream and eat the ice-cream just outside the store on a bench.)
Because within this study only the top-1 selected idea was presented to the teachers, the rater scores for the second and third idea were excluded from further analyses.

Interrater agreement.

The ICCs between the trained two raters indicated excellent agreement between the two raters for all three measures for the science task: .81 for originality, .79 for completeness, and .82 for practicality (consistency; single measures; Cicchetti, 1994). For the social task, the ICCs were .87 for originality, .73 for completeness, and .89 for practicality, indicating good to excellent agreement. Ratings were averaged across raters to gain the absolute trained rater scores.

Teacher assessments of CPS

In line with the 5-point Likert scale rating scheme for raters, a teacher rating scheme was developed (Table 1) and applied by the teachers to gain CPS assessments of the teachers.

Table 1
Teacher CPS Rating Scheme for the Originality, Completeness and Practicality Indicators.

	Originality	Completeness	Practicality
Level 1	The idea is not original at all. It is very predictable and commonly known.	The idea is not complete at all. The idea does not include any steps or completely ignores the problem.	The idea is not practical at all. It is not possible at all in practice.
Level 2	The idea is somewhat original. The idea is mostly commonly known, but presented in a different way.	The idea is somewhat complete. The idea describes a rough or vague idea	The idea is somewhat practical. It is very difficult to translate the idea into practice.
Level 3	The idea is reasonably original. It contains familiar elements, but something has been added that makes it different from other ideas.	The idea is reasonably complete. Elements of the idea address the problem, how is not completely clear.	The idea is reasonably practical. It takes several steps to put the idea into practice, but it is possible.
Level 4	The idea is mostly original. Elements are combined in a unique way.	The idea is mostly complete. Multiple steps are described in the idea.	The idea is mostly practical. It only takes a single step to put the idea into practice.
Level 5	The idea is very original. It's out-if-the-box, unique, resourceful and imaginative.	The idea is very complete. It contains all the steps required to arrive at a solution.	The idea is very practical. The idea can be put into practice right away.

Absolute and relative scores.
The raw scores given by the teachers and the trained raters formed the absolute assessments of students CPS abilities. For the relative assessments, rank order scores were calculated. To construct rank order scores, the raw teacher scores and raw rater scores were separately class mean centered for each separate class.

Academic achievement
Scores from two standard Dutch achievement tests were used as a measure for academic achievement. The test for math ability and reading comprehension were selected because they give a broad image of a student’s academic achievement and were administrated at every school in our sample.

We calculated learning efficiency percentage scores for the widely used cito tests (a standardized test commonly used and administrated twice a year in Dutch elementary education; Janssen et al., 2010; Tomesen et al., 2019) for mathematical ability and reading comprehension by dividing the student's test score by the national mean test scores for grade (as provided by the test developer) and multiplying this by 100. This means that students that scored exactly on average for their age received a learning efficiency percentage of 100%. This gives us an indication of how a student performed compared to other students from the same grade and makes comparison across age groups possible. One school (two classes, N = 46) administrated a different standardized test (i.e., the Boom test; Van Vugt et al., 2019) to measure students' mathematical ability. For this test, learning efficiency percentages calculated in a similar way were provided by the test developer.

Procedure

Data collected for this study were part of a larger CPS data collection effort. The institutional ethical review board approved the study protocol in March 2018 (nr.: FETC16-058). The data collection took place in April, May and June 2018. Data were collected in two sessions at school in the regular classrooms, during school days. Data were collected by the first author and a trained research assistant using an administration protocol. Both sessions lasted approximately 45 minutes.

In session 1, students completed the two CPS tasks (one from the social and one from the science domain, in a random order) and a short questionnaire on how they experienced the exercises. During the first session, teachers completed a background questionnaire. In session 2, students completed a third CPS exercise (from the entrepreneurial domain), a short questionnaire on how they perceived the tasks and a divergent thinking task. This third CPS exercise, the short questionnaire on how the tasks were experienced, the entrepreneurial task, and the divergent thinking task were not part of the current study. During the second session, teachers rated students' creative ideas collected in the first session using Qualtrics survey software. This survey started with an instructional video for teachers of about 10 minutes duration, including definitions of CPS, the goals of the study and a description of the CPS tasks involved. Importantly, the rating schemes were presented and shortly discussed with three examples. The top 1 ideas students of the teacher's class came up with on the science and social task were digitally presented as pictures to the teachers in the next step of the survey. The teachers were first asked to rate the completeness of the ideas, before rating the practicality and originality of the ideas.

Analysis

Analyses were conducted in IBM SPSS Statistics version 26 and Mplus version 8.5 (Muthén & Muthén, 2012). To answer our research question and determine whether teachers can assess CPS, we first studied to what extent teacher ratings of students CPS abilities (originality, practicality, completeness) correspond with trained rater assessments of these skills. For this purpose, the interrater reliability was assessed between the teachers and the trained raters for the corresponding skills. Intraclass correlations (one-way mixed, single measures) were calculated in SPSS, for the absolute and relative scores on both tasks. Again, Cicchetti's (1994) rules of thumb were used to classify the ICCs.

Next, regression models were run in Mplus to determine whether trained rater ratings of students CPS abilities predicted teacher assessments of these skills. To screen for potential assessment biases, students' academic achievement and gender were included in a second step. Standardized regression coefficients and explained variance (R^2) were calculated in both steps and classified using Cohen's (1988) rules of thumb. Changes in explained variance were reviewed to determine whether academic achievement and gender explained any additional variance ($\Delta R^2 > .05$) in the CPS scores. Again, the regression models were run for both the absolute and relative scores.

As is often the case in creativity studies, highly original ideas were relatively scarce and practical ideas were relatively common within the current sample. As a result, the distribution of these variables was non-normal. To take this non-normality and the clustering of students in classes into account, the functions type = complex and type = twolevel were applied in Mplus as methods of analysis. The Robust Maximum Likelihood (MLR) estimator applied with these methods is robust to non-normality and non-independence (Byrne, 2013).

5.3 Results

Missing data

Data of 10 students was missing due to unscorable tasks, so data of 610 students could ultimately be used in the analysis. Partly missing data of the students (6.6%) occurred because students were absent/ill during (a part of) the first or second session and because one teacher was unable to score the originality of the science task due to technical problems. This resulted in slightly different sample sizes for every analysis (Table 2; Table 3). A small group of students did not define a top-3, or their top-3 could not be interpreted. This was the case for 11 science exercises (1.95%) and 9 social exercises (1.59%). As pilot study results indicated that students tended to be quite able to define a top 3 (see Chapter 2), we defined a top-3 for these exercises ourselves. The same two raters scored the top-3 ideas. Academic achievement data was missing for mathematics for 15 students (2.51%) and for reading comprehension for 21 students (3.53%). The MLR estimator used for the analyses enabled us to estimate the model using all the available data.

Table 2
Descriptives of Teacher and Trained Rater Absolute and Relative CPS Scores and Students' Math and Reading Comprehension Learning Efficiency Percentages.

	Teacher Rated CPS		Trained Rater Rated CPS	
	M (SD)	Min/Max	M (SD)	Min/Max
Absolute Scores				
Science Originality	2.34 (1.30)	1/5	2.09 (1.05)	1/5
Science Completeness	2.92 (1.25)	1/5	3.24 (0.81)	1/5
Science Practicality	3.87 (1.36)	1/5	3.24 (0.81)	1/5
Social Originality	1.99 (1.14)	1/5	2.03 (1.03)	1/5
Social Completeness	2.96 (1.23)	1/5	3.30 (0.68)	1/5
Social Practicality	3.67 (1.33)	1/5	4.27 (0.85)	1/5
Relative Scores				
Science Originality	0 (1.13)	-2.23/3.44	0 (0.96)	-2.02/3.38
Science Completeness	0 (1.01)	-2.23/3.44	0 (0.76)	-2.42/2.35
Science Practicality	0 (1.23)	-3.62/2.35	0 (0.77)	-2.42/2.35
Social Originality	0 (1.02)	-3.62/2.35	0 (0.97)	-1.74/2.98
Social Completeness	0 (0.86)	-3.81/3.04	0 (0.65)	-2.40/2.08
Social Practicality	0 (0.86)	-3.54/2.47	0 (0.81)	-3.00/1.35
	M (SD)	**Min/Max**		
Math LE%	98.51(11.70)	51.09/135.63		
Reading comprehension LE%	98.51(11.70)	51.09/135.63		

Alignment between teachers and trained raters

Interrater reliability

To evaluate the alignment between teacher assessments and trained rater assessments of CPS, the interrater reliability (Intraclass correlation; one-way mixed single measures) between teachers and trained raters was calculated for both the absolute and relative (class mean centered) CPS scores of the science and social CPS task (Table 3).

With regard to the absolute assessment of CPS (i.e., the raw CPS scores), the interrater reliability for the originality and practicality indicators ranged from .41 to .55 for both tasks, indicating fair agreement between the teachers and the trained raters (Cicchetti, 1994). For the completeness indicator, the interrater reliability was .38 and .22 for the science and social tasks respectively. Thus, for the absolute completeness assessments, the agreement between the teachers and the trained raters was poor (Cicchetti, 1994). For the relative assessment scores (i.e., the class mean centered scores), the inter-rater reliability ranged from .41 to .54 for all three indicators (originality, completeness, & practicality) for both tasks, indicating fair agreement between the teachers and the trained raters (Cicchetti, 1994).

Table 3

Inter-Rater Reliability and 95% Confidence Interval of Absolute and Relative Teacher Ratings with the Corresponding Absolute and Relative Trained Rater Ratings.

	N	Trained Rater Ratings Absolute	Relative
Teacher Rating Science Task			
Originality	545	.50 [.43, .56]	.50 [.43, .56]
Completeness	571	.38 [.30, .44]	.53 [.47, .59]
Practicality	571	.43 [.36, .50]	.50 [.43, .56]
Teacher Rating Social Task			
Originality	570	.55 [.48, .60]	.54 [.48, .59]
Completeness	570	.22 [.14, .30]	.41 [.34, .48]
Practicality	570	.41 [.33, .47]	.49 [.42, .55]

Note: ICC's < .40 are classified as 'poor'; .40 - .59 as 'fair'; .60 - .74 as 'good' and > .74 as 'excellent'; Cicchetti, 1994.

Trained rater assessments predicting teacher assessments

The first step of the regression analysis on the absolute assessment scores revealed that trained rater assessments significantly predicted teacher assessments of the corresponding skill (β's ranging from .27 to .55 for all aspects, for both tasks; $p < .001$; Table 4). The explained variance (R^2) for CPS originality was .26 and .30 for the science and social task respectively, indicating a substantial share of the variance in absolute teacher scores was explained by the absolute trained raters scores (Cohen, 1988). For CPS completeness, the explained variance (R^2) was .17 for the science task and .07 for the social task, indicating only a small to moderate part of the variance in absolute teacher scores could be explained with the absolute scores of trained raters (Cohen, 1988). The explained variance (R^2) for CPS practicality was .23 and .20 for the science and social task respectively, indicating a moderate share of the variance in absolute teacher scores was explained by the absolute trained raters scores (Cohen, 1988).

Also for the relative scores, trained rater assessments significantly predicted teacher assessments of the corresponding skill (β's ranging from .43 to .55 for all aspects and both tasks; $p < .001$; Table 5). For the CPS completeness scores on the social task, the explained variance was moderate ($R^2 = .19$). For the CPS completeness scores on the science task and the originality and practicality scores on both tasks, the amount of variance explained was substantial (R^2 ranging from .26 to .31).

Potential assessment biases

Students' academic achievement (standardized math and reading comprehension test scores) and gender did not explain an additional relevant share of the variance in teacher assessments of CPS ($\Delta R^2 \leq .05$). For the absolute assessment of CPS, only reading comprehension positively predicted teachers' originality scores on the social task ($\beta = .08$, $p < .05$). However, this predictor explained only a minimal amount of additional variance ($\Delta R^2 < .01$). For the relative assessment of CPS, reading comprehension positively predicted teachers' practicality scores on the science task ($\beta = .07, p < .05$) and gender predicted teachers' completeness scores on the social task, in favor of girls ($\beta = .10, p < .05$). Again, these predictors explained only a small amount of additional variance ($\Delta R^2 = .03$ and .01 respectively).

Table 4
Results of Regression Models for the Absolute CPS Scores: Trained Rater Ratings Predicting Teachers' Ratings (Model 1) and Trained Rater Ratings, Academic Achievement, and Gender Predicting Teachers' Ratings (Model 2).

	Model 1		Model 2					
	β_{TR}	R^2	β_{TR}	β_{MA}	β_{RC}	β_{GE}	R^2	ΔR^2
Science								
Originality	.51**	.26	.50**	.01	.06	.02	.26	<.01
Completeness	.41**	.17	.42**	-.01	-.02	.01	.17	<.01
Practicality	.48**	.23	.52**	-.07	.09	.05	.28	.05
Social								
Originality	.55**	.30	.54**	.01	.08*	-.01	.30	<.01
Completeness	.27**	.07	.26**	.02	.02	.05	.07	<.01
Practicality	.45**	.20	.45**	-.03	.04	.05	.21	.01

*Note: ** p < .001; * p < .05; TR = trained rater ratings; MA = math test scores; RC = reading comprehension test scores, GE = gender (0 = boy, 1 = girl).*

Table 5
Results of Regression Models for the Relative CPS Scores: Trained Rater Ratings Predicting Teachers' CPS Ratings (Model 1) and Trained Rater Ratings, Academic Achievement, and Gender Predicting Teachers' CPS Ratings (Model 2).

	Model 1		Model 2					
	β_{TR}	R^2	β_{TR}	β_{MA}	β_{RC}	β_{GE}	R^2	ΔR^2
Science								
Originality	.51**	.26	.51**	-.03	.05	.05	.27	.01
Completeness	.55**	.31	.55**	-.01	.02	.04	.31	<.01
Practicality	.55**	.31	.57**	-.08	.07*	.04	.34	.03
Social								
Originality	.54**	.29	.53**	-.02	.09	-.01	.29	<.01
Completeness	.43**	.19	.42**	-.00	.03	.10*	.20	.01
Practicality	.51**	.26	.51**	-.06	.01	.07	.27	.01

*Note: ** p < .001; * p < .05; TR = trained rater ratings; MA = math test scores; RC = reading comprehension test scores, GE = gender (0 = boy, 1 = girl).*

Differences between teachers

For the absolute scores in particular, a large part of the variance in the teachers' CPS assessments remained unexplained. In order to gain insight in whether this variance might be due to differences between teachers we reviewed the variances at the class level and compared the intraclass correlations (ICC(1)) of the teachers and the trained raters. As might be expected, for the trained raters, there was little variance at the class level (ICC(1) ranging from .02 to .14, see Table 6). However, for teachers, more variability in the CPS assessments at the class level was found (ICCs ranging from .14 to .48, see Table 6). Especially the intraclass correlations for the assessment of completeness were large (ICC(1) science = .31; ICC(1) social = .48).

Table 6

Variance (s) Within and Between Classes and ICCs for the Teacher and Trained Rater Assessments.

	Teachers			Trained Raters		
	s within	*s* between	ICC(1)	*s* within	*s* between	ICC(1)
Science task						
Originality	1.34**	0.34**	.20	0.97**	0.13*	.12
Completeness	1.07**	0.47**	.31	0.61**	0.04*	.06
Practicality	1.58**	0.26**	.14	0.63**	0.10*	.14
Social task						
Originality	1.08**	0.23**	.18	0.98**	0.08*	.08
Completeness	0.78**	0.71**	.48	0.45**	0.01	.02
Practicality	1.31**	0.43**	.25	0.68**	0.04*	.06

*Note: ** $p < .001$; * $p < .05$*

5.4 Discussion

The goal of this study was to determine whether elementary school teachers can assess students' CPS abilities. For this purpose, we investigated whether teacher assessments of CPS abilities (i.e., the ability to produce original, complete, and practical ideas) were aligned with trained rater assessments, for both the absolute (i.e., raw) CPS scores and the relative (i.e., class mean centered) scores of students. The main finding of this study was that teachers are fairly able to assess students' CPS abilities, as long as they can focus on comparing students within their class (i.e., using relative scores). Applying an absolute norm appeared to be doable when originality scores were assigned, but the absolute assessment of practicality and, in particular, completeness, appeared to be more difficult.

Teacher and trained rater agreement

For the originality indicator, both the interrater reliabilities between the teachers and the trained raters and the results of the regression analyses showed that the teacher assessments were fairly aligned with the assessments of the trained raters. Regarding the originality of the ideas, it did not matter what task was assessed and whether an absolute 'standard' or a relative 'rank order' CPS norm was applied, the teachers and trained raters agreed rather well with each other in all cases.

This means that this study sheds a more positive light on teachers' abilities to assess indicators of creativity: in contrast to earlier studies (Beghetto et al., 2011; Gralewski & Karwowski, 2013; Urhahne, 2011) teachers in this study were quite well able to assess the originality of ideas that students come up with.

For the completeness indicator, the interrater reliabilities revealed that teacher assessments were only well aligned with the trained rater assessment when relative scores were used. For the absolute assessment of completeness (i.e., the raw scores), the agreement between the teachers and the trained raters was rather poor. Especially applying a completeness 'norm' in the assessment of the social task seemed to be difficult for teachers. That this absolute assessment of completeness was a challenge for teachers is illustrated by the intraclass correlations calculated in the additional analysis: especially for the assessments of completeness the variances at the class level were large. The fair agreement for the relative scores indicates that teachers do see which ideas are more or less complete, but that they differed in how they applied the 'norm' of the completeness scale. In other words: some teachers may be more rigid and therefore – structurally – assign lower scores, whereas some teachers may be more lenient in their scores.

For the practicality indicator, no notable differences across tasks were found. Furthermore, the teacher assessments for this indicator tended to be slightly better aligned with the assessments of trained raters when these scores were used to rank students in a classroom. In other words, also for the assessment of the practicality indicator, teachers tended to struggle with applying the practicality 'norm'. They seemed however well able to assess the practicality of the ideas in a relative way.

A possible explanation for the finding that the agreement between teachers and trained raters was higher for the originality indicator than for the other two, is the so-called 'originality bias'. This means that teachers tend to relate creativity to imagination and the constitution of original products, but to struggle to include criteria related to the usefulness or appropriateness of the ideas that are captured with the completeness and practicality indicators (Beghetto, 2010; Mullet et al., 2016). Either way, it seems clear that teachers do see which ideas are more complete or practical, but they simply seem to differ in how they apply the scale for these two indicators. These differences may have occurred because teachers differ in the degree of detail they demand from their students or they may have diverging opinions on what a completely 'practical' solution is and thus which ideas should receive the full five points for practicality. Nonetheless, when the aim is to assess CPS in an absolute way, more training, for instance with more assessment examples, is necessary for teachers to adequately understand the different levels of the completeness and practicality rating procedures and to apply them accordingly.

Within this study, potential biases of teachers were investigated by analyzing whether students' gender and academic achievement predicted teacher assessments of CPS, apart from the trained rater assessments. Only the reading comprehension test scores significantly predicted teachers' absolute originality assessment scores on the social task and the teachers' relative practicality scores on the science task. In contrast to the findings of Gralewski and Karwowski (2013) and in line with Sommer and colleagues (2008), gender only predicted teachers' relative completeness assessments on the social task, in favor of the girls. However, academic achievement and gender did not explain an additional, relevant share of the variation in the teacher assessments. We therefore conclude that, in general, the teachers in this study did not seem to be biased by gender and academic achievement in their CPS assessments. These findings indicate that teachers, at least in the Dutch elementary school context, do not mistake creativity for efficiency of school functioning, as was suggested in earlier research (e.g., Gralewski & Karwowski, 2013). They also do not seem to notably favor boys or girls in their assessments of creativity (Gralewski & Karwowski, 2013; Sommer et al., 2008). These findings are encouraging because they suggest that teachers do not have to eliminate certain biases towards specific students first: they can immediately focus on validly assessing CPS in all of their students.

Overall, as some argue, it is almost impossible to develop a 'standard' or norm for measures of creativity (Harris, 2016). The assessment of CPS is influenced by many factors, such as students' motivation to solve the problem, previous experiences with the problem situation and whether students think they are able to solve the problem (Reiter-Palmon et al., 2009). Consequently, every CPS task tends to produce different absolute creativity scores. This seems to be mirrored by our results. Applying a creativity 'norm' might therefore be very difficult for teachers. Applying a relative norm seems therefore more appropriate and, probably even more important, certainly more feasible for teachers. In addition to that, for most purposes – especially for everyday teaching and (formative) assessment purposes in the classroom - standardized absolute scores might be less important (Baer & Kaufman, 2019). In fact, earlier research showed that providing high school students with feedback information on relative performance had a strong positive effect on students' eventual performance (Azmat & Iriberri, 2010). The relative CPS assessment of the teachers may as such be used in the classroom to provide students with feedback and this may, potentially, help students to improve their CPS abilities. Creativity requires a safe environment without pressure or negative feedback (Amabile et al., 2002; Redifer et al., 2021). Beghetto (2005) therefore recommends to apply a mastery goal structure in the classroom that emphasizes goal related, relative feedback and focuses on students' individual development. The teacher rubric applied in this study may assist in setting these goals with students and may help to provide them with relative performance feedback to enhance their CPS abilities.

Limitations

Few studies have explicitly studied teacher assessments of students' CPS abilities. This study therefore provided us with highly needed information about the reliability of teachers' CPS assessments and gave us insight in how assessments of CPS could be applied in elementary education. Within this study, teacher assessments of CPS were compared to trained rater assessments of the corresponding skills. Although the same indicators and five-point scales were used by both the teachers and the trained raters to make comparison possible, the rating schemes used by the trained raters were first adopted to fit the teachers' needs. As an example, a few sentences were shortened or rewritten and fewer examples were given to make the rating scheme less complicated. This might have compromised teachers' full understanding of the CPS indicators. Although these modification were important to enable teachers to assess CPS, findings should still be interpreted with care.

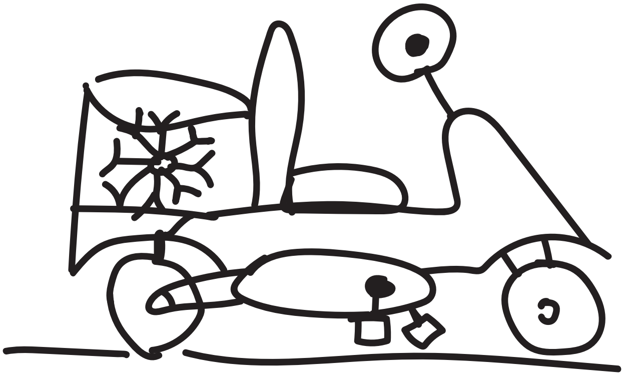
Within this study, 26 teachers and their classes participated. Although the large number of students allowed us to study CPS in greater detail at the student level, the number of teachers is too small to reliably assess actual differences between teachers and to predict this variability with teacher or classroom variables. The large variances at the class level (see also Table 6) do, nonetheless, indicate that teachers potentially vary greatly in how they assess CPS. This might not only be due to a different application of the 'standard', but might also be explained by other teacher characteristics. For example, teacher characteristics that were found to be associated with creativity evaluation skills were for instance teachers' own creative abilities, personality characteristics, age, and gender (Benedek et al., 2016). In a future study, a comparable design could be applied with more teachers rating the same set of ideas stemming from CPS tasks, to gain even more insight in the assessment of CPS.

Conclusions

In conclusion, teachers' CPS assessments may be used within a class to determine and follow elementary students' CPS abilities, but can – at this moment – not be used to directly compare students' CPS scores across classes. Within this study, a CPS task, a teacher rating scheme and a short, digital training were developed to assess CPS in elementary education. By following this training and by assessing CPS with the teacher ratings scheme, teachers may gain a better image of the CPS abilities of students and may consequently be better able to nurture these CPS abilities in students. To determine the actual consequences of the assessment of CPS, it is important to study what teachers do with these scores. Future studies could therefore address whether teachers are better able to differentiate their instructions with these CPS scores, to support the CPS needs of a variety of learners (e.g., Treffinger et al., 2008). Apart from that, it would be relevant to study how the assessment of CPS shapes teachers' perceptions of creativity and creative students (e.g., Anderson et al., 2022; Kettler et al., 2018). Either way, more insight into how CPS can be assessed and nurtured accordingly may assist in establishing a CPS habit in students that will help them to cope with the unknown problems of tomorrow.

6

GENERAL DISCUSSION



"Mount a freezer on the back of your bike."

CPS abilities are regarded as necessary to function in our rapidly changing society (Craft, 2011). When I was a teacher, I noticed that many colleagues struggled with embedding CPS in their everyday teaching practices, probably because they missed the resources to promote CPS in their students (cf. Davies et al., 2014). Furthermore, my fellow teachers experienced that the increasing weight of standardized tests leave little room for CPS in the classroom (cf. Black & Wiliam, 1998; Solomon, 2009). Within this PhD project, a CPS assessment procedure was developed that can assist teachers in assessing CPS in the classroom. Because assessment tends to drive learning (van der Vleuten et al., 2000), this assessment procedure and its resources (such as the structured CPS assessment task developed within this project) may help teachers to nurture CPS in students in the future. Before any CPS assessment procedure can be transferred to education, it is important to better understand CPS in primary school students. This project had, therefore, two main goals: (1) to gain insight into the nature of CPS in primary school students, and (2) to formulate implications for the assessment of CPS.

6.1 Summary of the Findings per Chapter

Chapter 2

Me and my colleagues first studied whether applying the CPS steps of fact finding (i.e., the exploration of knowledge) and problem finding (i.e., defining the problem) contributed to eventual CPS outcomes. Therefore, in the first study, it was examined whether these two processes of the 'understanding the challenge' element of the CPS model (Isaksen et al., 2011) contributed to the eventual creativity of the generated ideas. Additionally, we studied whether solution-finding (i.e., selecting the most creative ideas) is doable for elementary school students and how they select the ideas they think are the most creative ones.

The results indicated that a successful exploration of knowledge and a high-quality problem definition were positively associated with students' ability to find more, and more original ideas. In addition, problem finding seemed to be positively associated with the completeness of ideas, whereas fact finding was not. The results also showed that, in general, primary school students are able to identify their most creative solutions. We did not find any indications that the students undervalued aspects of originality, completeness, or practicality when defining their top 3. If anything, primary school students tended to overvalue rather than disregard completeness. The findings were rather similar across the two CPS tasks that were used.

Chapter 3

In Chapter 3, we described how researchers can conduct generalizability studies to evaluate creativity assessment procedures (cf. Brennan, 2010). Within this chapter, a generalizability study is described which was conducted to find out whether the CPS scores obtained with the CPS assessment procedure could be generalized across tasks. Overall, this study gave insight into how raters and tasks impact CPS scores and how generalizations differ for absolute and relative assessments of CPS.

Within this generalizability study CPS was assessed using two structured tasks (a science 'melting ice pops' task and a social 'distraction in class' problem task) and two raters (same sample as applied in Chapter 2). The Generalizability study revealed that if only this set of tasks and raters is used to assess CPS abilities, students' scores will also reflect other factors than solely student CPS

performance, in other words, these scores will be biased.
The seemingly large impact of the rater on the originality scores suggested that for CPS, it is quite difficult to get raters completely in line on which ideas are original and which ideas are not, which is comparable to the scoring of divergent thinking tasks (Forthmann et al., 2017). A large share of the variance was attributed to the tasks for the practicality scores. This indicates that the ability to come up with practical ideas when doing CPS highly depends on the specific problem presented to the students. These findings imply that it is important to offer multiple, different tasks to students to gain a better picture of a student's CPS abilities.

Based on the results of the generalizability study, a series of D-studies were executed. Results showed that for absolute decisions about a student's ability to produce original ideas, (too) many CPS tasks and raters are necessary to reach a generalizability of .70. This shows that it is very difficult to generalize the CPS performance across tasks in an absolute (i.e., 'norm based') way. However, when the aim is to make relative decisions (i.e., determine how students perform compared to other students from the same class), only three to five tasks rated by two raters can yield reliable CPS scores. This means that the CPS assessment procedure developed for this project may be used to assess CPS in a more formative way, describing students' relative performance.

Chapter 4
In the studies described in Chapter 4, the goal was to gain further insight into the nature of CPS processes in primary school students in terms of both the CPS process (as described by the model of Isaksen and colleagues, 2011) and the product (the set of creative ideas to solve a problem, scored on the four indicators) of primary school students.

The results of the studies largely matched with both CPS theory (Isaksen et al., 2011) and results from previous studies in other age groups. In the think-aloud study of the CPS process (study 4.1), every student (N = 13) showed CPS processes belonging to each of the four main elements of the CPS model (understanding the challenge, generating ideas, preparing for action, planning your approach; Isaksen et al., 2011). In the large-scale quantitative study (study 4.2) on the CPS product, the good fit of the measurement model and the correlations between the indicators demonstrated that the four CPS indicators (fluency, originality, completeness, and practicality) were distinguishable in the ideas of primary school students. The relations of these CPS indicators largely matched relations found in other age groups. The analysis of the structural model showed that CPS and divergent thinking are - also in primary school students – related concepts. In line with the meta-analysis of Gajda and colleagues (2017), small but significant relations between academic achievement and the CPS indicators originality and completeness were found. Therefore, the first conclusion was that the CPS model can be applied in the primary school context. The second conclusion was that CPS ability, as found in adults, is also already present in this young age group.

Chapter 5
In the study described in Chapter 5, we studied whether elementary school teachers are able to assess students' CPS abilities. We investigated whether teacher assessments of students' CPS (i.e., the ability to produce original, complete, and practical ideas) are aligned with trained rater assessments. Potential biases of teachers were included by studying whether students' gender and academic achievement predicted teacher assessments of CPS, apart from the trained rater assessments. Both the absolute CPS scores (i.e., 'norm based' scores) and the relative ordering of

students based on their CPS task performance, were analyzed to determine whether – and in what way - teachers can assess students' CPS.

The results of this study indicated that applying an absolute norm appeared to be only feasible for the originality indicator. When the originality of ideas was assessed, it did not seem to matter what task was assessed and whether an absolute 'standard' or a relative 'rank order' CPS norm was applied. In all cases, teachers and trained raters agreed rather well with each other. The teachers' and trained raters' relative CPS assessments tended to be quite well aligned for all three indicators (i.e., originality, completeness, and practicality). Additionally, the results indicated that teachers were not notably biased by students' gender or academic achievement. The main finding of this study was that the teachers could assess students' CPS, as long as they could apply relative scoring and, in this way, focus on comparing students within a class.

6.2 The Nature of CPS in Primary School Students

Within this section, I will elucidate what the implications of the studies, that I conducted with my colleagues, are for our understanding of the nature of CPS in primary school students. I will first discuss how CPS as studied in this thesis, relates to divergent thinking and academic achievement. Furthermore, I will discuss how CPS in primary school students as studied for this thesis corresponds with CPS in adults and how teachers could assess CPS. I will end with discussing how our studies could contribute to our understanding of the task- or domain-specific nature of CPS.

1. CPS and divergent thinking
Traditionally, creativity is assessed with divergent thinking tasks within and outside education. Within divergent thinking tasks, students are, for instance, asked to think of alternative uses for an object like a brick, a paperclip, or a toothbrush. However, as described in the introduction, a CPS process not only entails a divergent thinking or idea generation element, but also includes elements making room for convergent thinking processes. These elements are also called fact-finding (i.e., the exploration of knowledge), problem-finding (i.e., defining the problem at stake), and solution-finding (i.e., the evaluation of ideas to identify the most creative ones; Isaksen et al., 2011).

In the study described in Chapter 2, we found that especially the exploration (i.e., explicit activation) of knowledge seemed to kickstart the idea-finding stage by activating more opportunities (cf. Isaksen et al., 2011; McCaffrey, 2016). However, fact-finding is seldom applied at the start of a divergent thinking task. More attention for this process of knowledge exploration at the start of a divergent thinking task may help students come up with more ideas (and consequently more creative ideas). In this study, solution-finding was applied using the top-scoring method: students selected their three most creative ideas. Findings showed that students were quite able to select creative ideas, and no aspect of creativity (i.e., originality, completeness, and practicality) tended to be undervalued. This implies that top-scoring may not only be applied in divergent thinking tasks (e.g., Benedek et al., 2013; Silvia et al., 2008), but also in, more comprehensive, CPS tasks.

The study described in Chapter 4 revealed that the correlation between CPS fluency and divergent thinking fluency was quite strong ($r = .50$). This indicates that when students are asked to generate solutions to solve a problem, (partly) similar processes are activated as when students are asked to generate ideas on a divergent thinking task. Within this project, both the CPS stimuli (the problems) and the divergent thinking stimuli (i.e., What can one can do with a toothbrush?) may be regarded as concrete: both stimuli stem from everyday practice. In some divergent thinking tasks, more abstract stimuli are applied (Forthmann et al., 2019; Reiter-Palmon et al., 2019). Future research could address whether the correrelations between CPS fluency and divergent thinking fluency remain this strong when stimuli become more abstract.

In the study described in Chapter 4, we also found that the correlations between CPS originality and divergent thinking originality and between CPS completeness and divergent thinking elaboration were significant, although effects were small (r =.16 and r = .27 respectively). As suggested in Chapter 4, this might be explained by the different ways of conceptualizing and measuring CPS and divergent thinking outcomes. For instance, for the CPS tasks, an idea was regarded as original when the trained raters (i.e., creativity researchers that have seen ideas of a larger group of students) regarded it as original. Within the divergent thinking task, a statistical infrequency approach was applied: an idea scored high on originality when no other student in the sample mentioned it. For the latter, this means that an eventual divergent thinking originality score depends on the sample a student is in. This may not be a problem. The definition of Plucker et al. (2004) mentions that creativity is always bound to a specific, social context. Researchers should, however, still be aware of the dependency of these originality scores using a statistical infrequency approach.

6

In sum, some similarities were found between CPS and divergent thinking, but also some differences were detected. I conclude that where similar processes and procedures are demanded and applied (e.g., fluency processes and the top-scoring procedure), CPS and divergent thinking resemble each other rather closely. When CPS and divergent thinking are conceptualized and, as a consequence, are measured differently, differences occur. Therefore, researchers studying creative processes should carefully select the processes under study and the measurements applied in their research design (e.g., Reiter-Palmon et al., 2019). If teachers intend to nurture creativity in the classroom, it is probably more relevant to offer CPS tasks to students, rather than only divergent thinking tasks. CPS tasks explicitly embed steps focusing on both divergent and convergent thinking and focus on solving problems that children may actually face in real life (Isaksen et al., 2011). In addition, in CPS tasks a direct link can be made with knowledge learned at school, which may help students to apply their knowledge in a more flexible way.

2. CPS and academic achievement

In the study described in Chapter 4, we related the aggregated CPS indicators to outcomes from common academic achievement tests measuring mathematical ability and reading comprehension. The correlations that were found were significant, but rather small. On the one hand, this indicates that the CPS assessment task does, as intended, measure something different than academic achievement and is, therefore, of additional value. On the other hand, stronger relations might be expected between CPS and achievement tests, when CPS not only draws on creativity, but also on domain knowledge (e.g., Isaksen et al., 2011). However, the domain knowledge demanded in the CPS task (i.e., the science, social, and entrepreneurial task) and the achievement tests (i.e., the math and reading comprehension tests) we used, was not the same. Gajda and colleagues (2017) therefore recommend to add domain-specific achievement measures to study the observed relationship between creativity and academic achievement. In future research, it would therefore be interesting to relate the CPS findings to other student outcomes from more similar domains, such as STEM grades or reports of students' social skills. Future research could also more thoroughly study how CPS and academic achievement could complement each other in the primary curriculum. Earlier research indicated that creativity and mathematics could be integrated in one program and that this does not hamper the, more general, development of mathematical performance (Schoevers et al., 2020b). Studying how CPS can be integrated in, for instance, science lessons, allows us to gain insight in the 'both/and approach' as suggested by Beghetto (2013) and to, more structurally, embed CPS in primary education.

3. Primary school students versus adults

Although CPS is often applied in adult samples in, for instance, the workplace (e.g., Myszkowski et al., 2015; Puccio et al., 2018), CPS in primary school students remains understudied. The findings from our studies gave us more insight in the nature of the CPS ability of this young age group. A finding from the study described in Chapter 2 was that primary school students, like adults, successfully applied top scoring and were, as such, able to select their most creative ideas (e.g., Benedek et al., 2013). In addition, the think-aloud study described in Chapter 4 showed that primary school students apply CPS as described by Treffinger, Isaksen and colleagues (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995). In the larger-scale quantitative study described in the same chapter, similar relations between the CPS indicators were detected, as were found in earlier studies with adults (e.g., Dumas & Dunbar, 2014; Reiter-Palmon et al., 2009). All of these findings suggest that CPS ability is already present in this young age group and that, potentially, it could be fostered in similar ways as in adults (e.g., Puccio et al., 2018; Ritter et al., 2020).

4. Assessing CPS

Teachers should be able to determine students' CPS abilities in order to nurture these abilities in the classroom (e.g., Benedek et al., 2016; Bolden et al., 2020). Earlier research showed that raters (in general) and, more specifically, teachers struggle to assess creativity and are often biased in their judgments (Beghetto et al., 2011; Gralewski & Karwowski, 2013; Urhahne, 2011). Especially determining the originality of ideas seems to be a challenge for raters (Forthmann et al., 2017). These findings correspond with our study in which we extensively trained raters (see Chapter 3): the larger impact of the rater on the originality scores suggests that raters diverge in their originality ratings. In the teacher study (see Chapter 5), teachers were minimally trained. Their assessments of originality were nonetheless quite well in line with the originality assessments of the trained raters. It is not evident that the agreement between trained raters and minimally trained teachers would improve further with increased training of the teachers. Assigning originality scores may always be somewhat intuitive, even when raters are trained, which could result in differences between raters.

The question is also who should decide what is creative or not. The debate about this question has been going on for some time (e.g., Kaufman & Baer, 2012; Kaufman et al., 2016b). In this debate, some authors argue that when the aim is to foster creativity, students should be able to judge whether their own ideas might be considered creative (Kaufman et al., 2016b; Urban & Urban, 2022). This form of creative self-evaluation is also called creative metacognition and has lately received more attention in the creativity literature (Kaufman et al., 2016b). The top-scoring method implemented in the CPS-task in the current project may be regarded as an application of creative metacognition, because students have to evaluate the ideas and define their three most creative ones (Benedek et al., 2013). To make even more room for creative metacognition in CPS, future studies could, more explicitly and deliberately, ask students to discuss and evaluate their ideas themselves and study how this judgment process shapes future CPS processes. Overall, I believe that, in the future, students can have an even more active role in this assessment process.

5. Task- or domain-specificity of CPS

Although determining the task- or domain-specificity of CPS was not a central goal of this project, and the research design does not allow me to draw strong conclusions about this, the findings can still be discussed in light of this topic. In the study described in Chapter 3, we conceptualised CPS as an ability that can be generalized across tasks. These tasks stemmed from different domains: the science, social and entrepreneurial domain. However, the large variability across tasks within students' CPS found in Chapter 3 - especially for the 'practicality indicator '– could be taken as evidence that CPS is mainly a task-specific or domain-specific ability. However, the study described in Chapter 4 sheds a different light on this topic. The findings of this study showed that the originality, completeness, and practicality indicators can be aggregated across tasks. In addition, the relations between these constructs were largely in line with those found in other studies. In other words, when we attempted to segment and explain the 'noise' in CPS scores (as was done in Chapter 3), the domain specific aspects of CPS came to light. When we intended to eliminate the 'noise' and studied overlap in scores across tasks, domain general aspects became apparent. This implies that, although it is important to practice CPS within a specific domain, data from different CPS tasks can be used to say something about students' general CPS abilities, or – in other words – that CPS may have domain-general characteristics.

Plucker and Beghetto (2004) argue that each component of their definition of creativity (i.e., the interplay between ability and process, the observable outcome or product, the novelty and usefulness, and the social context) probably has both domain-specific and domain-general characteristics. On the one hand, creative abilities are task-specific: They can only be practiced within a domain due to the specific nature of knowledge that is needed to come to creative ideas in a problem (cf. Schoevers et al., 2020a; Willemsen et al., 2020). On the other hand, research showed that general cognitive abilities and creative problem-solving processes can be applied across domains (cf. Qian et al., 2019; Willemsen et al., 2020). Plucker and Beghetto (2004) conclude that especially for education, the distinction between general and specific approaches does not matter and that the development of creativity in students is hindered by too much generality or too much specificity. They recommend a hybrid position between these two and advise teachers to expose students to a variety of contexts in which they can apply their creativity in a search for an optimal interaction of both ability and context.

This perspective is in line with the results of our studies: when the CPS ability of students is measured in different contexts, variability across CPS tasks will – inevitably – occur, as it is most likely to occur across tasks within a domain (Reiter-Palmon et al., 2009). It is, however, still worthwhile to practice CPS across tasks and domains because (a) the task- or domain-general CPS procedures are, in this way, continuously practiced, and (b) students are given the chance to be creative in different contexts and to find the context they can be most creative in. Therefore, I conclude that CPS is – most likely – both task- or domain general and task- or domain-specific. Stimulating teachers to take this hybrid position when developing their teaching may help to properly transfer CPS to educational practice.

6.3 Implications for the Assessment of CPS

Apart from these more general insights, this project also leads to immediate implications for the assessment of CPS.

1. The steps of the structured assessment task 'work'
In Chapter 2, we evaluated whether the steps of the structured CPS task contributed to eventual CPS-outcomes. Here, we found that the exploration of knowledge and defining the problem seemed beneficial to find more and more original ideas. In addition, we also found that students were able to select their most creative ideas. For this reason, I conclude that the steps of the structured CPS task were relevant, i.e., 'worked', and could be applied accordingly in future assessments of CPS.

However, it is important to note that the results also suggest that defining the problem could serve to find more complete ideas and that the exploration of knowledge may even hamper it. Earlier research showed that task instructions impact creativity, in the sense that the 'focus' of the task instructions is mirrored in eventual outcomes (Di Mascio et al., 2018; Nusbaum et al., 2014). Attention to completeness in the task instructions (e.g., asking students to be both original and complete in the idea generation stage) could, in future assessments of CPS, potentially counterbalance this lack of focus.

2. The CPS outcomes are a product of a CPS process
The think-aloud study reported in Chapter 4, indicated that the CPS model provided by Treffinger, Isaksen and colleagues (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) matched the CPS behaviors of the students on the structured CPS task. Although it should be noted that a direct comparison was not possible and that the think-aloud sample was small, I do believe that the results from the think-aloud study indicate that the theoretical CPS-processes are actually engaged by the students in the CPS-assessment tasks. Because the same tasks are used to assess outcomes, I therefore conclude that the CPS outcomes are a product of a CPS process.

3. The CPS indicators can be applied to assess CPS in primary school students
The results of the study described in Chapter 4 showed that the four CPS indicators (i.e., fluency, originality, completeness, and practicality) could be constructed with the data of the CPS tasks. In addition, the relations between these indicators and divergent thinking outcomes were largely in line with earlier creativity research in different and mostly older samples (e.g., Dumas & Dunbar, 2014; Reiter-Palmon et al., 2009). This implies that the CPS indicators can be used to describe the CPS outcomes in primary school students.

4. Focus on the relative assessment of CPS
The findings from Chapter 5 indicated that teachers struggle to apply norm-based assessments of CPS, especially for the completeness and practicality indicators. These findings were in line with the findings from Chapter 3: generalizing CPS ability across tasks in an absolute way proved not be feasible in practice. However, when the goal is to determine a student's relative CPS-performance, only a few tasks may already give a reliable image. Applying a 'norm' for creativity is difficult (e.g., Guilford, 1950; Said-Metwaly et al., 2017), and the summative, standardized assessment of creativity in education has been criticized for this reason (e.g., Harris, 2016). Because creativity

interacts with other factors, such as motivation and perceived task characteristics, every creativity task tends to yield different absolute creativity scores (e.g., Reiter-Palmon et al., 2009). Therefore, applying a norm could wrongfully undervalue a student's creativity, while this student may be rather creative on a particular task compared to other students of the same class. A relative and formative approach to the assessment of CPS may therefore be favorable in education.

6.4 Directions for Future Research

Within research on the assessment of creativity, generalizability and other forms of reliability are perhaps the most difficult aspects of validity to provide support for (Barbot, 2019). This is reflected in the generalizability study described in Chapter 3: generalizing scores from a CPS task to a student's absolute CPS ability tended to be almost impossible. Apart from the more relative and formative approach for educational purposes suggested earlier, I recommend doing more generalizability studies to investigate how the generalizability of CPS assessments can be further improved.

Although the primary school students participating in the studies were young (i.e., about 9 to 12 years old; grade 4 to 6), formal schooling starts even earlier. Earlier research suggested that CPS may already be present in even younger children (e.g., Kim et al., 2019). It would therefore be interesting to explore whether and how CPS can be assessed and fostered from, for instance, kindergarten onwards.

The study described in Chapter 5 does give us a first idea of what the assessment of CPS by teachers could look like. To determine the actual consequences of the assessment of CPS, it is important to study what teachers do with these scores. Future research could therefore address how teachers' CPS assessments shape CPS teaching practices and how this, in turn, impacts students' development of CPS abilities. More specific research questions could be addressed as well. For instance: Are teachers able to differentiate their CPS instructions to support the needs of a variety of learners? Or: How does the assessment of CPS shape teachers' perceptions of students?

6.5 Limitations

The studies described in this thesis were, for a large part, conducted in the classroom, which contributes to the ecological validity of this research. This also inevitably incurred some limitations. These limitations were described in detail in the discussion sections of the respective chapters. Limitations that concern the thesis as a whole are described here.

First, the structured CPS task was developed for everyday educational practice. For this purpose, the CPS model was simplified, and some steps that are part of the complete CPS process were omitted in the task design. For instance, students were not instructed to 'construct opportunities' at the start of the CPS tasks and were not asked to 'build acceptance' for their solutions (see Isaksen et al., 2011 for the full model). Although the think-aloud study described in Chapter 4 showed that some of the students did engage in these processes, explicitly instructing students to apply these extra steps might have altered the results.

Although the decision study described in Chapter 3 gives an indication of how many tasks should be administrated to generalize results, this scenario is still hypothetical and based on the tasks included in the design (Brennan, 2010). When more tasks are constructed for the CPS assessment procedure, it is important to, once again, study the generalizability of the new design.

For the first two studies described in this thesis, data were collected in small groups outside the classroom. For the larger data collection that was part of the studies described in Chapters 4 and 5, data were (for practical and ecological validity reasons) collected in the classroom with the whole class. The tasks were, however, still completed individually. The administrators asked students frequently to work alone, but because creativity tends to benefit from an open class climate without competition and time-pressure (Davies et al., 2013; Eisenberg & Thompson, 2011), the administrators did not ‘punish’ students when they, accidently, talked or expressed themselves. As a consequence, it could not completely be prevented that students copied ideas from each other, which may have impacted the fluency and originality scores in particular.

The ultimate goal of working on CPS tasks with students is establishing a ‘creative habit for life’. This means that, in the end, students should be able to apply CPS in everyday life without a specific pre-described structure. Even though I think it is necessary to structure the CPS task for instructional and assessment purposes first, it is important to note that the results of these studies do not reflect a ‘free’ CPS process. Working towards a more flexible CPS process in the classroom is important to develop this habit and to make room for individual differences in creative problem-solving style (Main et al., 2019; Treffinger et al., 2008).

6.6 Recommendations for Teachers

Of course, I cannot finish this dissertation without formulating recommendations for my former colleagues that play a key role in developing CPS in students: the teachers. I hope these recommendations will be useful for them when implementing CPS tasks in the classroom.

Recommendation 1: Make room for fact-finding, problem-finding, and solution-finding
Explicitly paying attention to the exploration of knowledge (i.e., fact-finding) and defining the problem at stake (i.e., problem-finding) may help students to connect CPS to knowledge learned at school and may help to increase the creativity of ideas. Furthermore, applying solution-finding in the classroom by discussing and evaluating the ideas students come up with may help to make room for creative metacognition (i.e., creative self-evaluation; Kaufman et al., 2016b). Furthermore, this may help students to recognize creativity in their own and someone else’s work in the future.

Recommendation 2: Offer CPS regularly in different contexts
An implication I extensively discussed in the section on domain- or task-specificity, is that I recommend teachers to engage students in CPS regularly in different tasks and domains (Plucker & Beghetto, 2004) or even in interdisciplinary ways. In this way, students can repeatedly practice the CPS processes and are, in addition, given the chance to explore CPS in different contexts.

Recommendation 3: Be aware of the originality bias
Creativity is often equated with originality (Beghetto, 2010; Mullet et al., 2016). This originality bias is not surprising given that originality is widely recognized as an essential factor of creativity.

However, originality by itself is not sufficient: an idea is regarded creative when it is both original and serves a particular purpose (i.e., is regarded useful or appropriate; Plucker et al., 2004). This means that when CPS is applied in the classroom, all indicators of CPS should be, explicitly, valued. In other words, my advice for teachers would be to evaluate the students' ideas on all four indicators: originality, fluency, completeness, and practicality.

Recommendation 4: Be aware of the Big-C bias
What is a creative person? Often, eminent scientists or artists like Einstein or Mozart come to mind when answering this question. This emphasis on eminent creativity, also called the Big-C bias, resulted in the belief that creativity is extremely rare and only matters at these most eminent levels (Beghetto, 2010). However, creativity should be regarded as a continuum: it ranges from mini-c (intrapersonal) and little-c (everyday) creativity, through Pro-C (non-eminent, professional) to eminent forms of Big-C creativity (Einstein and Mozart). Although a CPS process may occasionally result in a Pro-C or Big-C innovation, teaching CPS primarily focuses on everyday creativity. As a consequence, the ideas resulting from a CPS process are most often creative at the intrapersonal or interpersonal (e.g., in the classroom) level. In this way, CPS can be regarded as a useful ability for everyone. For this reason, I recommend that teachers focus on fostering CPS ability in every student.

Recommendation 5: Apply a mastery goal orientation
The findings of the studies in this thesis support a relative and formative approach to assessing CPS ability in the classroom. This implies that I recommend teachers to focus on the relative standing of students instead of focusing on the application of a standardized norm. However, an open focus on ranking students within a classroom and competition between classmates should be omitted: this may undermine rather than foster creativity. For this reason, in line with Beghetto (2005), I recommend a mastery goal structure that emphasizes goal-related feedback and focuses on students' individual development. The rubric applied in our research (see Chapter 5) may assist in setting these goals and may help to provide feedback to students to develop themselves about the CPS indicators.

6.7 General conclusion

With this thesis, I aimed to gain insight into CPS ability in primary school students and how it may be assessed. In order to do so, a CPS assessment procedure was developed. The studies described in this thesis provided insights into how CPS relates to divergent thinking and academic achievement. In addition, the results show that primary school students can do CPS and that a relative, formative approach to the assessment of CPS ability may be most appropriate in education.

Multiple implications for the assessment of CPS ability were described in this final chapter. Although the tasks and rating procedures of our CPS assessment procedures were developed with care, the most sensitive aspect of the CPS assessment procedure probably still remains the generalizability of CPS scores. Guilford did already note in 1950 that "reliabilities of tests of creative abilities and of creative criteria will probably be generally low. There are ways of meeting such difficulties, however. We should not permit them to force us to keep foot outside the domain." (p. 445). Conducting more generalizability studies in creativity research, as was suggested in Chapter 3, could provide us with detailed information on how the generalizability of creativity assessments, and this CPS assessment procedure in particular, can be further improved in the future.

The research that was conducted within this project primarily focused on the nature and the assessment of CPS in primary school students and not (yet) on how CPS can be nurtured in primary school students. The structured task developed in this project does seem fit to foster CPS in primary school students. Although more research is needed to study the (long-term) benefits of doing CPS with students with this task, regularly doing CPS in the classroom may help teachers foster CPS abilities in the classroom (cf. Kim et al., 2019). Although I recommend, as was mentioned earlier in Chapter 2, to scaffold the CPS stages first, teachers could take a more flexible approach later on. This may make room for individual CPS learning goals and trajectories and may enable students to find their own CPS style (Main et al., 2019; Treffinger et al., 2008).

Due to the limitations described above, findings should be interpreted with care. Nevertheless, I also formulated five recommendations for teachers that they can apply in the classroom immediately. I hope that the findings of this thesis and these recommendations help teachers better understand creativity and CPS in particular. The ultimate goal of the studies described in this thesis was to contribute to the assessment of CPS in the primary school context and, in this way, enable teachers to embed CPS in the classroom. The CPS assessment procedure that was developed for this purpose may provide teachers with evidence-based resources to assess and nurture CPS in the classroom. By regularly applying the full CPS model in a variety of contexts, while being aware of potential biases and by applying a mastery goal orientation, teachers may be able to formatively assess and foster CPS abilities in students. In this way, teachers can significantly contribute to developing a creative habit for life, which may help students cope with the unknown problems of tomorrow.

NEDERLANDSE SAMENVATTING

"Use a Formula 1 bike."

Nederlandse samenvatting

Onze maatschappij verandert constant. We beseffen ons steeds meer dat kennis alleen niet voldoende is, innovatie en creativiteit worden steeds belangrijker. Het is dan ook zinvol om kinderen al vanaf jonge leeftijd te leren om problemen creatief aan te pakken (Craft, 2011). Dit levert niet alleen originele en bruikbare oplossingen op, maar het leert kinderen ook om flexibel om te gaan met de kennis die ze hebben. Daarnaast kunnen ze hun creatieve vaardigheden verder ontwikkelen. Hier ligt dan ook een taak voor het onderwijs: leraren kunnen zo al vanaf jonge leeftijd het creatief probleemoplossen bij hun leerlingen stimuleren.

Creatief probleemoplossen wordt in de literatuur ook wel creative problem solving genoemd (CPS) en is een vorm van creativiteit waarbij leerlingen zowel vakkennis als creatieve vaardigheden inzetten om alledaagse problemen op te lossen (Craft, 2011; Isaksen & Treffinger, 2004; Okuda et al., 1991). CPS kan in de klas geoefend worden in drie stappen. Eerst kunnen leerlingen gestimuleerd worden om een probleem te begrijpen (understanding the challenge, Isaksen et al., 2011) door de kennis die ze al hebben over het probleem te verkennen (fact finding) en het precieze probleem af te bakenen (problem finding). Vervolgens kan aan ze gevraagd worden om meerdere creatieve ideeën te genereren om het probleem op te lossen (ook wel generating ideas of idea finding). Tot slot kunnen leerlingen gestimuleerd worden om na te denken over de stap naar de praktijk (preparing for action) en dus om hun ideeën te evalueren en de meest originele, maar haalbare ideeën te selecteren om in praktijk te brengen (solution finding). Hierdoor kunnen we ze leren om creativiteit ook te gaan herkennen (Isaksen et al., 2011; Sophonhiranrak et al., 2015). Door naast deze stappen aandacht te hebben voor hoe het proces loopt (ook wel planning your approach, bijv. Wanneer kan je naar een volgende stap of moet je weer even terug?) kunnen we ze daarnaast leren een CPS-proces te monitoren en flexibel toe te passen.

Maar hoe bepaal je dan of een leerling goed is in creatief probleemoplossen? Naast dat er gekeken kan worden naar hoeveel ideeën een leerling genereert, kan de creativiteit van de ideeën vastgesteld worden door te bepalen hoe uniek ideeën zijn, hoe uitgewerkt ideeën zijn en in welke mate ze uiteindelijk ook uitvoerbaar zijn in de praktijk. Kortom, de CPS-vaardigheden van een leerlingen kunnen bepaald worden door te kijken naar vloeiendheid, originaliteit, compleetheid en praktische bruikbaarheid.

Hoewel leraren het belangrijk vinden aandacht te schenken aan het ontwikkelen van creatieve vaardigheden bij leerlingen (Kasirer & Shnitzer-Meirovich, 2021), worstelen ze met het implementeren van CPS in hun onderwijs (Kim et al., 2019). Ze gebruiken nu bijvoorbeeld met name klassieke divergent denken taken. In deze taken worden leerlingen bijvoorbeeld gevraagd om meerdere en creatieve gebruiksmogelijkheden te bedenken voor een paperclip of een tandenborstel. Divergent denken omvat dan ook het vermogen om meerdere, verschillende ideeën te genereren (De Vink et al., 2020). Convergent denken daarentegen omvat selectie- en evaluatievaardigheden, met als doel die ideeën te vinden die het meest kansrijk, origineel en bruikbaar zijn. In een compleet creatief proces zou er naast divergent denken expliciet ruimte moeten zijn voor convergent denken (Cropley, 2006). CPS behelst zowel divergent als convergent denken en past daarmee bij wat er geleerd wordt op school, omdat zowel domeinkennis als algemene creatieve processen gevraagd worden (Burnard et al., 2006; Poddiakov, 2011). Middels CPS zou creativiteit dus een plek kunnen krijgen binnen vakken die aangeboden worden op school.

We weten echter nog niet precies hoe CPS er bij leerlingen in het basisonderwijs uitziet. Er is wel al onderzoek gedaan naar CPS bij volwassenen, maar het onderzoek bij deze jonge doelgroep is schaars. Zo weten we bijvoorbeeld niet hoe leerlingen CPS toepassen en hoe je dan vervolgens kan bepalen of dit goed lukt. Het is van belang dat we meer zicht krijgen op CPS bij leerlingen in het basisonderwijs, om het uiteindelijk ook gericht te kunnen stimuleren.

Om leerlingen gericht te kunnen helpen met het ontwikkelen van hun CPS-vaardigheden, moeten leraren het ook kunnen beoordelen. Helaas is het bepalen van de CPS-vaardigheden van leerlingen überhaupt al lastig en een nog grotere uitdaging voor leraren in het bijzonder (zie bijv. Beghetto et al., 2011; Gralewski & Karwowski, 2013; Urhahne, 2011). Creativiteit wordt beïnvloed door diverse factoren, zoals taakkenmerken (zoals bijvoorbeeld hoe complex of relevant iemand een taak vindt; Reiter-Palmon et al., 2009) en domeinspecifieke creatieve vaardigheden (bijv. verschillen in domeinspecifieke kennis of vaardigheden; Barbot et al., 2016b). Als gevolg hiervan presteren leerlingen verschillend op creativiteitstaken en is het niveau van een leerling met een enkele taak lastig te bepalen. Daarnaast zijn er ook weinig tools beschikbaar om CPS in de klas op een efficiënte manier te beoordelen en te onderwijzen (Davies et al., 2014).

Het onderzoek dat ik in dit proefschrift behandel heeft dan ook twee doelen:
(1) meer inzicht krijgen in hoe CPS er bij basisschoolleerlingen uit ziet en
(2) aanbevelingen formuleren voor het onderwijs om CPS op een betrouwbare manier te beoordelen.

Binnen dit promotietraject heb ik daarom, samen met mijn team, een CPS assessmentprocedure ontwikkeld die docenten kan helpen bij het beoordelen van CPS in de klas. Deze assessmentprocedure bevat een gestructureerde CPS-taak, drie probleemsituaties uit het natuurwetenschappelijke, sociale en ondernemersdomein en richtlijnen voor het beoordelen van CPS.

S

Om dit te verder te onderzoeken heb ik vier studies gedaan die in de hoofdstukken 2 t/m 5 worden beschreven.

Hoofdstuk 2

In het onderzoek dat beschreven wordt in hoofdstuk 2 onderzocht ik eerst of het toepassen van de CPS-stappen fact finding (d.w.z. het verkennen van kennis) en problem finding (d.w.z. het definiëren van het probleem) bijdragen aan de uiteindelijke CPS-uitkomsten vloeiendheid, originaliteit, compleetheid en praktische bruikbaarheid. Daarnaast onderzochten we of de basisschoolleerlingen (N = 137; M leeftijd = 10.50) solution finding (d.w.z. het selecteren van de meest creatieve ideeën) kunnen toepassen en hoe zij de ideeën selecteren die volgens hen het meest creatief zijn. Dit hebben we onderzocht met twee verschillende taken: een natuurwetenschappelijke taak en een sociale taak.

De resultaten toonden aan dat een succesvolle verkenning van kennis en een goede probleemstelling positief samenhingen met het vermogen om meer en originelere ideeën te vinden. Bovendien leek de kwaliteit van de probleemstelling positief samen te hangen met de compleetheid van ideeën. Uit de resultaten bleek ook dat basisschoolleerlingen in staat zijn om hun meest creatieve oplossingen te selecteren. We vonden geen aanwijzingen dat leerlingen aspecten van originaliteit, compleetheid of praktische bruikbaarheid negeren bij het bepalen van hun meest creatieve ideeën, al leken ze de compleetheid van een idee het zwaarst mee te wegen. Tussen de twee taken werden weinig verschillen gevonden. De belangrijkste conclusies waren dan ook dat aandacht voor het verkennen van kennis en het bepalen van het probleem zinvol kan zijn voor de uiteindelijke creativiteit van de ideeën. Daarnaast lijken leerlingen in het basisonderwijs een creatieve oplossing best goed te kunnen herkennen.

Hoofdstuk 3

In hoofdstuk 3 hebben we aan de hand van dezelfde data als hoofdstuk 2 beschreven hoe onderzoekers Generalizability Theory kunnen toepassen binnen creativiteitsonderzoek (cf. Brennan, 2010). Met Generalizability Theory kan onderzocht worden hoe een assessment procedure verbeterd kan worden door bijvoorbeeld meer taken of beoordelingen toe te voegen. In dit hoofdstuk onderzochten we dan ook eerst of CPS-scores over taken heen kunnen worden gegeneraliseerd en daarmee of we iets kunnen zeggen over de algehele CPS-vaardigheid van leerlingen. Hierin werd een onderscheid gemaakt tussen absolute scores en relatieve scores, oftewel scores van leerlingen ten opzichte van andere leerlingen binnen de eigen klas.

De resultaten wezen uit dat als alleen een set van twee taken en twee beoordelaars zou worden gebruikt om de CPS-vaardigheden van leerlingen in kaart te brengen, de scores geen goed beeld van de CPS-vaardigheden van een leerling zouden geven.

Verder bleek dat het zeer moeilijk is om CPS-vaardigheden over taken heen op een absolute (d.w.z. normatieve) manier te meten. Hier zouden simpelweg te veel taken en beoordelaars voor nodig zijn. Wanneer het echter voldoende is om zicht te hebben op hoe leerlingen presteren in vergelijking met andere leerlingen uit dezelfde klas, kunnen slechts drie tot vijf door twee beoordelaars beoordeelde taken betrouwbare CPS-scores opleveren. Dit betekent dat de in dit project ontwikkelde CPS-assessmentprocedure wel kan worden gebruikt om CPS op een meer relatieve manier, binnen de klas, te beoordelen.

Hoofdstuk 4

De studies beschreven in hoofdstuk 4 hadden als doel meer inzicht te genereren in hoe leerlingen in het basisonderwijs creatief problemen oplossen. Hierbij hebben we zowel gekeken naar het CPS-proces (zoals beschreven door het model van Isaksen en collega's, 2011) als het CPS-product (de uiteindelijke ideeën om het probleem op te lossen, beoordeeld met de vier indicatoren) van basisschoolleerlingen.

De resultaten van deze studies kwamen grotendeels overeen met zowel de CPS-theorie als met resultaten uit eerdere studies met andere leeftijdsgroepen. In de think-aloud studie van het CPS-proces (zie 4.1) pasten alle leerlingen (N = 13) CPS-processen toe die horen bij elk van de vier hoofdelementen van het CPS-model (understanding the challenge, generating ideas, preparing for action, planning your approach; Isaksen et al., 2011). Uit de tweede, grootschaligere kwantitatieve studie (N = 594; zie 4.2) naar het CPS-product bleek dat de CPS-indicatoren (vloeiendheid, originaliteit, compleetheid en praktische bruikbaarheid) bruikbaar waren om de ideeën van de leerlingen, bedacht tijdens het maken van drie CPS-taken, te beoordelen. Daarnaast bleek dat relaties tussen deze CPS-indicatoren grotendeels overeenkwamen met relaties die eerder in andere leeftijdsgroepen werden gevonden. Ook bleek dat CPS en divergent denken (Bedenk zoveel mogelijk verschillende en originele dingen die je kan doen met een tandenborstel?) - ook bij basisschoolleerlingen - verwante concepten zijn. In lijn met de meta-analyse van Gajda en collega's (2017) werden kleine, maar significante relaties gevonden tussen academische prestaties (gemeten met bekende gestandaardiseerde testen voor rekenen en begrijpend lezen) en de CPS-indicatoren originaliteit en compleetheid. De eerste conclusie was dan ook dat het CPS-model toepasbaar is in de basisschoolcontext. De tweede conclusie was dat CPS-vaardigheid, zoals gevonden bij volwassenen, ook al aanwezig is bij basischoolleerlingen.

Hoofdstuk 5

In de studie beschreven in hoofdstuk 5 onderzochten we of de beoordelingen van leraren van de CPS-vaardigheden van leerlingen (d.w.z. het vermogen om originele, complete en praktische ideeën te produceren) overeenkomen met de beoordelingen van getrainde beoordelaars. Potentiële vooroordelen van leraren werden meegenomen door te onderzoeken of het geslacht en de academische prestatie van leerlingen de beoordelingen van docenten ook voorspelden. Zowel de absolute CPS-scores als de relatieve ordening van leerlingen (d.w.z. hoe ze binnen een klas beoordeeld worden, ten opzichte van elkaar) werden geanalyseerd om te bepalen of - en op welke manier - leraren de CPS-vaardigheden van leerlingen kunnen beoordelen.

De resultaten van deze studie gaven aan dat als leraren CPS-vaardigheden van leerlingen beoordelen, het toepassen van een absolute ‘norm’ lastig is bij het bepalen van de compleetheid en praktische bruikbaarheid van ideeën. De absolute beoordelingen voor deze twee indicatoren van de leraren verschilden simpelweg te veel van de getrainde beoordelaars. Bij het beoordelen van de originaliteit van ideeën ging dit beter: hier leek het niet uit te maken welke taak werd beoordeeld en of een absolute of een relatieve norm werd toegepast. In alle gevallen waren de leraren en de getrainde beoordelaars het behoorlijk met elkaar eens. De relatieve CPS-beoordelingen van leraren en de getrainde beoordelaars kwamen wel behoorlijk overeen voor alle drie de indicatoren. Bovendien bleek uit de resultaten dat leraren zich niet noemenswaardig lieten leiden door het geslacht of de academische prestaties van hun leerlingen. De belangrijkste bevinding van deze studie was daarom ook dat leraren CPS-vaardigheden kunnen beoordelen, mits de beoordelingen op een relatieve manier worden gebruikt en zich dus richten op het vergelijken van leerlingen binnen een klas.

Discussie

Het eerste doel van mijn onderzoek was om meer inzicht te krijgen in hoe CPS er bij basisschoolleerlingen uitziet. Hieronder volgen daarom de belangrijkste inzichten die de vier studies ons hebben opgeleverd.

CPS, Divergent denken taken en academische prestaties

Binnen dit onderzoek werden enkele overeenkomsten gevonden tussen CPS uitkomsten van basisschoolleerlingen en uitkomsten van traditionele divergent denken taken, maar ook enkele verschillen. Waar vergelijkbare processen werden gevraagd van leerlingen en vergelijkbare beoordelingsprocedures werden toegepast, bleken CPS en divergent denken vrij sterk op elkaar lijken. Zo werd er bijvoorbeeld een sterke correlatie gevonden tussen vloeiendheid binnen de CPS taak en vloeiendheid binnen de divergent denken taak. Wanneer CPS en divergent denken echter anders worden geconceptualiseerd en als gevolg daarvan anders worden gemeten, treden er verschillen op. Het is dan ook van belang dat onderzoekers de te bestuderen creativiteitsprocessen en de toegepaste metingen zorgvuldig selecteren (zie bijv. Reiter-Palmon et al., 2019).

Binnen dit onderzoek werden uitkomsten van CPS taken ook gerelateerd aan uitkomsten van schoolse, gestandaardiseerde kennis en vaardigheidstoetsen. De gevonden verbanden waren positief en statistisch significant, maar wel vrij klein. Enerzijds wijst dit erop dat de CPS taak (zoals bedoeld) iets anders meet dan de toetsen en daarom van toegevoegde waarde is binnen het curriculum. Anderzijds zouden sterkere relaties verwacht kunnen worden tussen CPS en scores op toetsen, omdat CPS niet alleen een beroep zou moeten doen op creativiteit, maar ook op domeinkennis (bv. Isaksen et al., 2011). Echter, de domeinkennis die gevraagd werd in de CPS taak (d.w.z. de natuurwetenschappelijke, sociale en ondernemerstaak) en de toetsen (d.w.z. de gestandaardiseerde toetsen voor rekenen en begrijpend lezen), was niet hetzelfde. Gajda en collega's (2017) raden daarom aan om domeinspecifieke kennismaten toe te voegen om de relaties tussen creativiteit en academische prestaties te bestuderen. In de toekomst zou het daarom interessant kunnen zijn om de CPS uitkomsten te relateren aan andere uitkomsten binnen het domein, zoals bijvoorbeeld aan uitkomsten van natuurwetenschappelijke kennistoetsen of aan gerapporteerde sociale vaardigheden.

Taak- of domeinspecificiteit van CPS

Hoewel het bepalen van de taak- of domeinspecificiteit van CPS geen centraal doel van dit project was, en de onderzoeksopzet mij niet toelaat hierover sterke conclusies te trekken, kunnen de bevindingen toch in het licht van dit onderwerp worden besproken. In het in hoofdstuk 3 beschreven onderzoek conceptualiseerden we CPS als een vaardigheid die over taken heen kan worden gegeneraliseerd. De grote variabiliteit tussen taken binnen leerlingen die in hoofdstuk 3 werd gevonden - vooral voor praktische bruikbaarheid - zou worden kunnen opgevat als bewijs dat CPS vooral een taak- of domeinspecifieke vaardigheid is. De in hoofdstuk 4 beschreven studie liet echter iets anders zien. De bevindingen van dit onderzoek toonden aan dat de indicatoren originaliteit, compleetheid en praktische bruikbaarheid over taken heen samengenomen kunnen worden. Bovendien kwamen de relaties tussen deze samengenomen constructen grotendeels overeen met de relaties die over het algemeen gevonden worden in creativiteitsonderzoek. Met andere woorden: toen we probeerden de 'ruis' in de CPS-scores te segmenteren en te verklaren (zoals in hoofdstuk 3), kwamen de domeinspecifieke aspecten van CPS naar voren. Toen we de 'ruis' wilden elimineren en juist de overlap in scores tussen de taken bestudeerden, kwamen de domeingenerieke aspecten aan het licht.

Plucker en Beghetto (2004) stellen dat iedere vorm van creativiteit waarschijnlijk zowel domeinspecifieke als domeingenerieke kenmerken heeft. Enerzijds zijn creatieve vermogens taak- of domeinspecifiek: Ze kunnen alleen binnen een domein geoefend worden vanwege specifieke domeinkennis en domeinvaardigheden die nodig zijn om tot creatieve ideeën te komen (Schoevers et al., 2020a; Willemsen et al., 2020). Anderzijds blijkt dat algemene cognitieve vaardigheden en creatieve probleemoplossingsprocessen over domeinen heen kunnen worden toegepast (Qian et al., 2019; Willemsen et al., 2020). Plucker en Beghetto (2004) beargumenteren dat voor het onderwijs het onderscheid tussen specifieke en generieke benaderingen er eigenlijk niet toe hoeft te doen en dat de ontwikkeling van creativiteit bij leerlingen wordt belemmerd door te veel specificiteit en algemeenheid. Zij bevelen een hybride positie tussen deze twee aan en adviseren leraren om leerlingen bloot te stellen aan diverse contexten waarbinnen zij hun creativiteit kunnen toepassen in een zoektocht naar een optimale interactie van creatief vermogen en context.

Dit perspectief komt dan ook overeen met de resultaten van onze studies. Daarom concludeer ik dat CPS - hoogstwaarschijnlijk - zowel taak- of domeinspecifiek als taak- of domeingeneriek is. Leraren stimuleren om een hybride positie in te nemen bij het toepassen van CPS in de praktijk is dan ook zinvol. Daarnaast blijft het de moeite waard om CPS in verschillende domeinen te oefenen, omdat zo (a) de taak- of domein-generieke CPS-procedures voortdurend worden geoefend en (b) leerlingen de kans krijgen om in verschillende contexten creatief te zijn.

Beoordeling van CPS

Docenten moeten de CPS-vaardigheden van leerlingen kunnen beoordelen om deze vaardigheden in de klas te kunnen stimuleren (Benedek et al., 2016; Bolden et al., 2020). Eerder onderzoek liet zien dat getrainde beoordelaars (in het algemeen) en, meer specifiek, leraren moeite hebben met het beoordelen van creativiteit en vaak bevooroordeeld zijn in hun oordeel (Beghetto et al., 2011; Gralewski & Karwowski, 2013; Urhahne, 2011). Vooral het bepalen van de originaliteit van ideeën lijkt een uitdaging te zijn (Forthmann et al., 2017). Deze bevindingen komen overeen met onze studie waarin we beoordelaars uitgebreid hebben getraind (zie hoofdstuk 3): de grote impact van de beoordelaar op de originaliteitsdata suggereert dat beoordelaars, ondanks training, nog altijd verschillen in hun beoordelingen. In de studie onder minimaal getrainde leraren (zie hoofdstuk 5) kwamen de originaliteitsbeoordelingen van leraren echter redelijk overeen met de beoordelingen van de getrainde beoordelaars. Het zou zo kunnen zijn dat het beoordelen van originaliteit altijd wat intuïtief is, of er nou weinig of veel training is toegepast. Meer onderzoek is echter nodig om dit verder uit te zoeken.

Uit ons onderzoek naar de beoordeling van CPS komen een aantal implicaties voort, die ik in de vorm van stellingen wil voorleggen:

1. De voorbereidende en afsluitende stappen van de gestructureerde CPS-taak zijn zinvol

In hoofdstuk 2 evalueerden we of de stappen van de gestructureerde CPS-taak bijdroegen aan uiteindelijke CPS-uitkomsten. Hier vonden we dat het verkennen van kennis en het definiëren van het probleem konden bijdragen aan het vinden van meer en originelere ideeën. Bovendien bleek dat de leerlingen hun meest creatieve ideeën konden selecteren. Daarom concludeer ik dat de stappen van de gestructureerde CPS-taak relevant zijn en op deze manier kunnen worden toegepast in toekomstige beoordelingen van CPS.

2. De CPS-uitkomsten zijn een product van een CPS-proces

De think-aloud studie die in hoofdstuk 4 werd gerapporteerd, gaf aan dat het CPS-model van Treffinger, Isaksen en collega's (Isaksen et al., 2011; Isaksen & Treffinger, 2004; Treffinger, 1995) overeenkwam met de CPS-gedragingen van de leerlingen op de CPS-taak. Deze resultaten lijken hiermee aan te geven dat de theoretische CPS-processen daadwerkelijk door de leerlingen worden uitgevoerd tijdens de taken. Omdat dezelfde taken worden gebruikt om de CPS-vaardigheden van leerlingen in kaart te brengen, concludeer ik daarom dat de CPS-uitkomsten een product zijn van een CPS-proces.

3. De CPS-indicatoren kunnen worden gebruikt om CPS-vaardigheden van basisschoolleerlingen in te schatten

De resultaten van het in hoofdstuk 4 beschreven onderzoek toonden aan dat de vier CPS-indicatoren (namelijk vloeiendheid, originaliteit, compleetheid en praktische bruikbaarheid) konden worden gebruikt om de CPS-vaardigheden van leerlingen in te schatten. Bovendien kwamen de relaties tussen deze indicatoren en de relaties met de uitkomsten van de divergent denken taak grotendeels overeen met eerder creativiteitsonderzoek met verschillende en veelal volwassen steekproeven (bijv. Dumas & Dunbar, 2014; Reiter-Palmon et al., 2009). Dit impliceert dat de CPS-indicatoren ook gebruikt kunnen worden om de CPS-vaardigheden van basisschoolleerlingen te beschrijven.

4. CPS kan, op een relatieve manier, beoordeeld worden

De bevindingen uit hoofdstuk 5 gaven aan dat leraren moeite hebben met het toepassen van absolute beoordelingen van CPS, met name voor de indicatoren compleetheid en praktische bruikbaarheid. Deze bevindingen kwamen overeen met de bevindingen uit hoofdstuk 3: het absoluut generaliseren van CPS-vaardigheid over taken heen bleek in de praktijk niet haalbaar. Wanneer echter het doel is om de relatieve positie van een leerling ten opzichte van andere leerlingen in de klas te bepalen, kunnen enkele taken al een betrouwbaar beeld geven. Het toepassen van een 'norm' voor creativiteit is moeilijk (e.g., Guilford, 1950; Said-Metwaly et al., 2017), en de summatieve, gestandaardiseerde beoordeling van creativiteit in het onderwijs wordt bekritiseerd (e.g., Harris, 2016). Omdat er zoveel andere factoren zoals motivatie en domeinspecifieke kennis meespelen, zou het toepassen van een norm zou de creativiteit van een leerling ten onrechte kunnen onderwaarderen. Een relatieve en daarmee meer formatieve manier van beoordelen van CPS is daarom passender binnen het onderwijs.

Beperkingen en suggesties voor vervolgonderzoek
De in dit proefschrift beschreven studies zijn voor een groot deel in de klas uitgevoerd, wat een bijdrage levert aan de ecologische validiteit van dit onderzoek. Dit zorgde, onvermijdelijk, ook voor enkele beperkingen.

Ten eerste werd de gestructureerde CPS-taak ontwikkeld voor de dagelijkse onderwijspraktijk. Hiervoor werd het CPS-model vereenvoudigd en werden sommige stappen die deel uitmaken van het volledige CPS-proces weggelaten in het ontwerp van de gestructureerde CPS-taak. Hoewel de in hoofdstuk 4 beschreven studie liet zien dat een deel van de leerlingen deze processen wel toepaste, zou het expliciet instrueren van leerlingen om deze extra stappen toe te passen andere resultaten op hebben kunnen leveren

Hoewel de in hoofdstuk 3 beschreven studie een indicatie geeft van het aantal taken dat moet worden toegepast in de klas om, op een betrouwbare manier, iets te kunnen zeggen over de CPS-vaardigheden van leerlingen, is dit scenario nog steeds hypothetisch en gebaseerd op de taken die in het ontwerp zijn opgenomen (Brennan, 2010). Wanneer meer taken zijn toegevoegd aan de assessmentprocedure, is het belangrijk om opnieuw de generaliseerbaarheid te onderzoeken.

De data voor deze onderzoeken is verzameld in groepen of met de hele klas tegelijk. Hierbij werd aan leerlingen gevraagd om alleen te werken. Omdat creativiteit doorgaans gebaat is bij een open klassenklimaat zonder competitie en tijdsdruk (Davies et al., 2013; Eisenberg & Thompson, 2011), 'straften' de onderzoekers leerlingen echter niet wanneer ze, per ongeluk, praatten of zich uitten. Daardoor kon niet volledig worden voorkomen dat leerlingen ideeën van elkaar kopieerden, wat met name de scores voor vloeiendheid en originaliteit kan hebben beïnvloed.

Hoewel de aan de studies deelnemende basisschoolleerlingen jong waren (d.w.z. ongeveer 9 tot 12 jaar oud; groep 5 t/m 8), start het basisonderwijs natuurlijk op nog jongere leeftijd. Eerder onderzoek suggereerde dat CPS mogelijk al aanwezig is bij kleuters (bijv. Kim et al., 2019). Het zou daarom interessant zijn om te onderzoeken of en hoe CPS kan worden beoordeeld en bevorderd in bijvoorbeeld de onderbouw van het basisonderwijs.

Het uiteindelijke doel van het werken aan CPS-taken met leerlingen is het creëren van een 'creative habit for life'. Dit betekent dat leerlingen uiteindelijk in staat moeten zijn CPS in het dagelijks leven toe te passen zonder een specifieke, vooraf beschreven structuur. Hoewel het noodzakelijk kan worden geacht om de CPS-taak eerst te structureren voor instructie- en beoordelingsdoeleinden, is het belangrijk om te noemen dat de resultaten van deze studies geen "vrij" CPS-proces weerspiegelen. Toewerken naar een meer flexibel CPS-proces in de klas is belangrijk om deze gewoonte te ontwikkelen en ruimte te maken voor individuele verschillen in creatieve probleemoplossingsstijl (Main et al., 2019; Treffinger et al., 2008).

De in hoofdstuk 5 beschreven studie geeft een eerste idee van hoe de beoordeling van CPS door leraren eruit zou kunnen zien. Toekomstig onderzoek zou zich vervolgens kunnen richten op de vraag hoe CPS bevordert kan worden in de klas en hoe de CPS-beoordelingen van leraren hierbij, op een zinvolle manier, gebruikt kunnen worden. De in dit project ontwikkelde gestructureerde taak lijkt geschikt om CPS bij basisschoolleerlingen te trainen. Meer inzicht is echter nodig in de voordelen van het doen van CPS met leerlingen met deze taak. Ook meer specifieke onderzoeksvragen zouden aan bod kunnen komen. Bijvoorbeeld: zijn leraren in staat hun CPS-instructies te differentiëren om de behoeften van verschillende leerlingen te ondersteunen? Of: hoe beïnvloedt de beoordeling van CPS de perceptie van leraren over (creatieve) leerlingen? Kortom: meer onderzoek naar hoe leraren CPS kunnen trainen in de klas is gewenst om het uiteindelijk structureel in te kunnen bedden in het basisonderwijs.

Aanbevelingen voor leraren

Ik kan dit proefschrift niet afsluiten zonder aanbevelingen te formuleren voor de mensen die een sleutelrol kunnen spelen bij het ontwikkelen van CPS bij leerlingen: de leraren.

Aanbeveling 1: Maak ruimte voor het exploreren van kennis, voor het definiëren van het probleem en voor het evalueren van creatieve ideeën

Expliciet aandacht besteden aan het verkennen van kennis (fact-finding, bijv. Wat weet je al over dit probleem?) en het definiëren waar het probleem echt om gaat (problem-finding) kan leerlingen helpen met het flexibel toepassen van kennis en kan daarnaast zorgen voor creatievere ideeën. Door meer ruimte te maken voor het evalueren van ideeën (solution-finding) leren kinderen daarnaast hun creativiteit beter te herkennen in hun eigen werk en in dat van anderen.

Aanbeveling 2: Bied CPS regelmatig aan in verschillende contexten

Een implicatie die ik heb besproken in de paragraaf over taak- of domeinspecificiteit is dat ik leraren adviseer om CPS regelmatig met leerlingen te oefenen, binnen verschillende domeinen. Zo kunnen leerlingen hun CPS-vaardigheden vergroten en krijgen ze bovendien de kans om te verkennen in welk vakgebied ze CPS goed of juist minder goed kunnen toepassen.

Aanbeveling 3: Wees je bewust van de 'originality bias'

Creativiteit wordt vaak gelijkgesteld aan originaliteit (Beghetto, 2010; Mullet et al., 2016). Deze 'originality bias' is niet verrassend: originaliteit wordt gezien als een essentiële factor van creativiteit. Originaliteit op zichzelf is echter maar een aspect van creativiteit. Een idee is pas creatief wanneer het én origineel is én een bepaald doel dient (d.w.z. nuttig of passend is; Plucker et al., 2004). Dit betekent dat wanneer CPS in de klas wordt toegepast, het zinvol is om de ideeën van de leerlingen te leggen naast alle vier de CPS-indicatoren: vloeiendheid, originaliteit, compleetheid en praktische bruikbaarheid.

Aanbeveling 4: Wees je bewust van de 'Big-C bias'

Wat is nu eigenlijk een creatief persoon? Vaak wordt hierbij gedacht aan eminente wetenschappers of kunstenaars zoals Einstein of Mozart. Deze nadruk op uitzonderlijke creativiteit, ook wel de Big-C bias genoemd, heeft geleid tot de overtuiging dat creativiteit uiterst zeldzaam is en er alleen toe doet op baanbrekende niveaus (Beghetto, 2010). Creativiteit moet echter worden beschouwd als een continuüm: het loopt van mini-c (intrapersoonlijke) en little-c (alledaagse) creativiteit, via Pro-C (professionele) tot eminente vormen van Big-C creativiteit (Einstein en Mozart). Hoewel een CPS-proces af en toe kan leiden tot een Pro-C of Big-C innovatie, richt het onderwijs in CPS zich vooral op alledaagse creativiteit. Als gevolg hiervan zijn ideeën die uit een CPS-proces voortvloeien meestal creatief op intra- of interpersoonlijk niveau, bijvoorbeeld binnen de klas of in de thuissituatie. CPS kan dus beschouwd worden als een nuttige vaardigheid voor iedereen. Daarom raad ik leraren aan zich te richten op het bevorderen van CPS-vaardigheden bij iedere leerling.

Aanbeveling 5: Pas beheersingsdoelen toe

Eerder gaf ik al aan dat het absoluut beoordelen van CPS lastig blijkt en dat daarom een relatieve norm binnen de klas bij het beoordelen van CPS het meest passend lijkt. Dit betekent dat leraren zich dus ook het beste kunnen richten op de relatieve ontwikkeling van leerlingen in plaats van op een – al dan niet gestandaardiseerde - norm. Een open focus op een rangorde en competitie tussen leerlingen binnen een klas is echter niet wenselijk: dit zal de creativiteit van leerlingen eerder ondermijnen dan bevorderen. Om deze reden adviseer ik, in lijn met Beghetto (2005), om beheersingsdoelen toe te passen. Hierbij wordt doelgerelateerde feedback toegepast, die zich richt op de individuele ontwikkeling van leerlingen. De in ons onderzoek toegepaste rubric (zie hoofdstuk 5) kan helpen bij het stellen van deze doelen en kan daarnaast leraren ondersteunen bij het geven van feedback aan leerlingen om zich verder te ontwikkelen.

Slotwoord

Het doel van dit PhD project was om beter zicht te krijgen op hoe CPS in het basisonderwijs toegepast en beoordeeld kan worden. Hiervoor werd een CPS-assessmentprocedure ontwikkeld en toegepast in de klas.

Ik hoop dat dit proefschrift en de geformuleerde aanbevelingen onderzoekers en leraren een beter beeld geven van creativiteit in het algemeen en CPS in het bijzonder. Het ultieme doel is dat het beoordelen van CPS, en daarmee de ontwikkeling ervan, een structurele plek krijgt in het basisonderwijs. De hiervoor ontwikkelde CPS-assessmentprocedure kan worden gezien als een evidence-based hulpmiddel om CPS in de klas te beoordelen en te oefenen. Op die manier kunnen leraren een belangrijke bijdrage leveren aan de ontwikkeling van de ‘creative habit for life’, die leerlingen kan helpen om te gaan met de onbekende problemen van morgen.

REFERENCES

"Simply re-freeze the ice pops in your freezer."

Altshuller, G. (1996). *And suddenly the inventor appeared: TRIZ, the Theory of Inventive Problem Solving*. Technical Innovation Center, Inc.
Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 153–158. https://doi.org/10.1037/aca0000231
Amabile, T. M. (1996). *Creativity in context*. Westview Press.
Amabile, T. M., Hadley, C. N., & Kramer, S. J. (2002). *Creativity under the gun.* Harvard Business Review, 80, 52–61
Amabile, T. M., & Pillemer, J. (2012). Perspectives on the social psychology of creativity. *The Journal of Creative Behavior, 46,* 3-15. https://doi.org/10.1002/jocb.001
Anderson, R. C., Katz-Buonincontro, J., Bousselot, T., Mattson, D., Beard, N., Land, J., & Livie, M. (2022). How am I a creative teacher? Beliefs, values, and affect for integrating creativity in the classroom. *Teaching and Teacher Education, 110*, 103583. https://doi.org/10.1016/j.tate.2021.103583
Arreola, N. J., & Reiter-Palmon, R. (2016). The effect of problem construction creativity on solution creativity across multiple everyday problems. *Psychology of Aesthetics, Creativity, and the Arts, 10*, 287–295. https://doi.org/10.1037/a0040389
Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics, 94*, 435-452. https://doi.org/10.1016/j.jpubeco.2010.04.001
Baer, J. (2012). Domain specificity and the limits of creativity theory. *The Journal of Creative Behavior, 46,* 16-29. https://doi.org/10.1002/jocb.002
Baer, J. & Garrett, T. (2017). Accountability, the common core, and creativity. In R.A. Beghetto & J.C. Kaufman (Eds.). *Nurturing creativity in the classroom.* (pp. 45-66). Sheridan Books.
Baer, J., & Kaufman, J. C. (2019). Assessing creativity with the consensual assessment technique. In I. Lebuda, & V. Glăveanu (Eds.). *The Palgrave handbook of social creativity research.* (pp. 27-37). Springer International. https://doi.org/10.1007/978-3-319-95498-1_3
Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?. *Metacognition and Learning, 3*, 39-58. https://doi.org/10.1007/s11409-007-9009-6
Barak, M. (2013). Impacts of learning inventive problem-solving principles: students' transition from systematic searching to heuristic problem solving. *Instructional Science, 41*, 657–679. https://doi.org/10.1007/s11251-012-9250-5
Barbot, B. (2019). Measuring creativity change and development. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 203–210. https://doi.org/10.1037/aca0000232
Barbot, B., Besançon, M., & Lubart, T. (2016a). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences, 52*, 178-187. https://doi.org/10.1016/j.lindif.2016.06.005
Barbot, B., & Lubart, T. (2012). Creative thinking in music: Its nature and assessment through musical exploratory behaviors. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 231–242. https://doi.org/10.1037/a0027307
Barbot, B., Lubart, T. I., & Besançon, M. (2016b). "Peaks, slumps, and bumps": Individual differences in the development of creativity in children and adolescents. *New Directions for Child and Adolescent Development, 151*, 33–45. https://doi.org/10.1002/cad.20152
Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the

scattered literature. *Genetic, Social, and General Psychology Monographs, 132*, 355-429. https://doi.org/10.3200/mono.132.4.355-430

Beghetto, R. A. (2005). Does assessment kill student creativity? *The Educational Forum, 69*, 254 –263.

Beghetto, R. A. (2010). Creativity in the classroom. In J. C. Kaufman & R. J. Sternberg (Eds.), *Cambridge handbook of creativity* (pp. 447– 463). Cambridge University Press.

Beghetto, R. A. (2013). *Killing ideas softly? The promise and perils of creativity in the classroom.* Information Age.

Beghetto, R. A., Kaufman, J. C., & Baxter, J. (2011). Answering the unexpected questions: Exploring the relationship between students' creative self-efficacy and teacher ratings of creativity. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 342-349. https://doi.org/10.1037/a0022834

Benedek, M., Muhlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts, 7*, 341–349. https://doi.org/10.1037/a0033644

Benedek, M., Nordtvedt, N., Jauk, E., Koschmieder, C., Pretsch, J., Krammer, G., & Neubauer, A. C. (2016). Assessment of creativity evaluation skills: A psychometric investigation in prospective teachers. *Thinking Skills and Creativity, 21*, 75–84. https://doi.org/10.1016/j.tsc.2016.05.007

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7–74.

Blair, C. S., & Mumford, M. D. (2007). Errors in idea evaluation: Preference for the unoriginal? *The Journal of Creative Behavior, 41*, 197–222. https://doi.org/10.1002/j.2162-6057.2007.tb01288.x

Bolden, B., DeLuca, C., Kukkonen, T., Roy, S., & Wearing, J. (2020). Assessment of Creativity in K-12 Education: A Scoping Review. *Review of Education, 8*, 343-376. https://doi.org/10.1002/rev3.3188

Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32*, 83-100. https://doi.org/10.1177/0265532214542994

Brennan, R. L. (2001). *Generalizability theory.* Springer.

Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in Education, 24*, 1-21. https://doi.org/10.1080/08957347.2011.532417

Brophy, D. R. (2001). Comparing the attributes, activities, and performance of divergent, convergent, and combination thinkers. *Creativity Research Journal, 13*, 439-455. https://doi.org/10.1207/S15326934CRJ1334_20

Burnard, P., Craft, A., & Cremin, T. (2006). Documenting 'possibility thinking': A journey of collaborative enquiry. *International Journal of Early Years Education, 14*, 243–262. https://doi.org/10.1080/09669760600880001

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Routledge.

Byrne, C. L., Shipman, A. S., & Mumford, M. D. (2010). The effects of forecasting on creative problem-solving: An experimental study. *Creativity Research Journal, 22*, 119-138. https://doi.org/10.1080/10400419.2010.481482

Ceh, S. M., Edelmann, C., Hofer, G., & Benedek, M. (2022). Assessing raters: What factors predict discernment in novice creativity raters?. *The Journal of Creative Behavior*, *56*, 41-54. https://doi.org/10.1002/jocb.515

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Laurence Erlbaum Associates

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381-398. https://doi.org/10.1037/1082-989X.12.4.381

Conti, R., Coon, H., & Amabile, T. M. (1996). Evidence to support the componential model of creativity: Secondary analyses of three studies. *Creativity Research Journal, 9*, 385-389. https://doi.org/10.1207/s15326934crj0904_9

Corazza, G. E. (2016). Potential originality and effectiveness: The dynamic definition of creativity. *Creativity Research Journal*, *28*, 258–267. https://doi.org/10.1080/10400419.2016.1195627

Craft, A. (2011). *Creativity and education futures.* Trentham Books.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* Wiley.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review*, *23*, 72-79. https://doi.org/10.1080/02783190009554069

Cropley, A. J. (2001). *Creativity in education & learning: A guide for teachers and educators.* Psychology Press.

Cropley, A. J. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*, 391–404. https://doi.org/10.1207/s15326934crj1803.

Crossley, J., Johnson, G., Booth, J., & Wade, W. (2011). Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Medical Education*, *45*, 560-569. https://doi.org/10.1111/j.1365-2923.2010.03913.x

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 159 – 166. https://doi.org/10.1037/ACA0000220

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments in education—A systematic literature review. *Thinking Skills and Creativity*, *8*, 80-91. https://doi.org/10.1016/j.tsc.2012.07.004

Davies, D., Jindal-Snape, D., Digby, R., Howe, A., Collier, C., & Hay, P. (2014). The roles and development needs of teachers to promote creativity: A systematic review of literature. *Teaching and Teacher Education*, *41*, 34-41. https://doi.org/10.1016/j.tate.2014.03.003

De Vink, I. C., Willemsen, R. H., Lazonder, A. W., & Kroesbergen, E. H. (2022). Creativity in mathematics performance: The role of divergent and convergent thinking. *British Journal of Educational Psychology*, *92*, e12459. https://doi.org/10.1111/bjep.12459

Di Mascio, R., Kalyuga, S., & Sweller, J. (2018). The effect of wording and placement of task instructions on problem-solving creativity. *The Journal of Creative Behavior*, *52*, 335-353. https://doi.org/10.1002/jocb.157

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56-67.

https://doi.org/10.1016/j.tsc.2014.09.003
Eisenberg, J., & Thompson, W. F. (2011). The effects of competition on improvisers' motivation, stress, and creative performance. *Creativity Research Journal*, *23*, 129-136. https://doi.org/10.1080/10400419.2011.571185
English, L. D. (1997). The development of fifth-grade children's problem-posing abilities. *Educational Studies in Mathematics*, *34*, 183–217. https://doi.org/10.1023/A:1002963618035
Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251. https://doi.org/10.1037/0033-295X.87.3.215
Finke, R. A., Ward, T. B., & Smith, S. M. (1992). Creative cognition: *Theory, research, and applications*. MITPress.
Forthmann, B., Holling, H., Zandi, N., Gerwig, A., Çelik, P., Storme, M., & Lubart, T. (2017). Missing creativity: The effect of cognitive workload on rater (dis-)agreement in subjective divergent-thinking scores. *Thinking Skills and Creativity*, *23*, 129-139. https://doi.org/10.1016/j.tsc.2016.12.005
Forthmann, B., Wilken, A., Doebler, P., & Holling, H. (2019). Strategy induction enhances creativity in figural divergent thinking. *The Journal of Creative Behavior*, *53*, 18-29. https://doi.org/10.1002/jocb.159
Gajda, A., Karwowski, M., & Beghetto, R. A. (2017). Creativity and academic achievement: A meta-analysis. *Journal of Educational Psychology*, *109*, 269–299. https://doi.org/10.1037/edu0000133
Gipps, C. V. (1994). Beyond testing: Towards a theory of educational assessment. Falmer.
Gralewski, J., & Karwowski, M. (2013). Polite girls and creative boys? Students' gender moderates accuracy of teachers' ratings of creativity. *Journal of Creative Behavior*, *47*, 290–304. https://doi.org/10.1002/jocb.36
Grohman, M., Wodniecka, Z., & Kłusak, M. (2006). Divergent thinking and evaluation skills: do they always go together? *Journal of Creative Behavior*, *40*, 125–145. https://doi.org/10.1002/j.2162-6057.2006.tb01269.x
Guilford, J.P. (1950). Creativity. *American Psychologist*, *5*, 444–454.
Guo, J., Ge, Y., & Pang, W. (2019). The underlying cognitive mechanisms of the rater effect in creativity assessment: The mediating role of perceived semantic distance. *Thinking Skills and Creativity*, *33*, 100572. https://doi.org/10.1016/j.tsc.2019.100572
Harms, M., Reiter-Palmon, R., & Derrick, D. C. (2018). The role of information search in creative problem solving. *Psychology of Aesthetics, Creativity, and the Arts*, *14*, 367-380. https://doi.org/10.1037/aca0000212
Harris, A. (2016). *Creativity and education*. Palgrave Macmillan.
Hass, R. W., Rivera, M., & Silvia, P. J. (2018). On the dependability and feasibility of layperson ratings of divergent thinking. *Frontiers in Psychology*, *9*, 1343. https://doi.org/10.3389/fpsyg.2018.01343
Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation procedures and a case for the generalizability study. *Educational Researcher*, *41*, 56-64. https://doi.org/10.3102/0013189X12437203
Hoijtink, H., van Kooten, P., & Hulsker, K. (2016). Why Bayesian psychologists should change the way they use the Bayes factor. *Multivariate Behavioral Research*, *51*, 2–10. https://doi.org/10.1080/00273171.2014.969364
Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative*

Health Research, *15*, 1277-1288. https://doi.org/10.1177/1049732305276687
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55. https://doi.org/10.1080/10705519909540118
Isaksen, S. G., Dorval, K. B., & Treffinger, D. J. (2011). *Creative approaches to problem solving: A framework for innovation and change.* Thousand Oaks.
Isaksen, S. G., & Treffinger, D. J. (2004). Celebrating 50 years of reflective practice: Versions of creative problem solving. *The Journal of Creative Behavior*, *38*, 75–101. https://doi.org/10.1002/j.2162-6057.2004.tb01234.x
Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 tot en met 8* [Scientific justification of the mathematics test for Grade 1 to Grade 6]. Cito.
Jiang, Z. (2018). Using linear mixed-effect model framework to estimate generalizability variance component in R: A lme4 package application. *Methodology*, *14*, 133-142. https://doi.org/10.1027/1614-2241/a000149
Jiang, Z., Raymond, M., Shi, D., & DiStefano, C. (2020). Using a linear mixed-effect model framework to estimate multivariate Generalizability Theory parameters in R. Behavior *Research Methods*, *52*, 2383-2393. https://doi.org/10.3758/s13428-020-01399-z
Kashani-Vahid, L., Afrooz, G., Shokoohi-Yekta, M., Kharrazi, K., & Ghobari, B. (2017). Can a creative interpersonal problem solving program improve creative thinking in gifted elementary students? *Thinking Skills and Creativity*, *24*, 175–185. https://doi.org/10.1016/j.tsc.2017.02.011
Kasirer, A., & Shnitzer-Meirovich, S. (2021). The perception of creativity and creative abilities among general education and special education teachers. *Thinking Skills and Creativity*, *40*, 100820. https://doi.org/10.1016/j.tsc.2021.100820
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
Kaufman, J. C. (2012). Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 298–308. https://doi.org/10.1037/a0029751
Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal*, *24*, 83-91. https://doi.org/10.1080/10400419.2012.649237
Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal, 20*, 171-178. https://doi.org/10.1080/10400410802059929
Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R., & Sinnett, S. (2013). Furious activity vs. understanding: How much expertise is needed to evaluate creative work? *Psychology of Aesthetics, Creativity, and the Arts*, *7*, 332-340. https://doi.org/10.1037/a0034809
Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology*, *13*, 1–12
Kaufman, J.C., Beghetto, R.A., & Dilley, A. (2016a). Understanding creativity in the schools. In A.A. Lipnevich, F. Preckel & R.D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century* (pp. 133–153). Springer.
Kaufman, J. C., Beghetto, R. A., & Watson, C. (2016b). Creative metacognition and self-ratings of creative performance: A 4-C perspective. *Learning and Individual Differences*, *51*, 394–399. https://doi.org/10.1016/j.lindif.2015.05.004

Kaufman, J. C., Cole, J. C., & Baer, J. (2009). The construct of creativity: Structural model for self-reported creativity ratings. *The Journal of Creative Behavior*, *43*, 119–134. https://doi.org/10.1002/j.2162-6057.2009.tb01310.x

Kaufman, J. C., Lee, J., Baer, J., & Lee, S. (2007). Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability. *Thinking Skills and Creativity*, *2*, 96-106. https://doi.org/10.1016/j.tsc.2007.04.002

Kettler, T., Lamb, K. N., Willerson, A., & Mullet, D. R. (2018). Teachers' perceptions of creativity in the classroom. *Creativity Research Journal*, *30*, 164-171. https://doi.org/10.1080/10400419.2018.1446503

Kim, S., Choe, I., & Kaufman, J. C. (2019). The development and evaluation of the effect of creative problem-solving program on young children's creativity and character. *Thinking Skills and Creativity*, *33*, 100590. https://doi.org/10.1016/j.tsc.2019.100590

Klugkist, I., van Wesel, F., & Bullens, J. (2011). Do we know what we test and do we test what we want to know? *International Journal of Behavioral Development*, *35*, 550–560. https://doi.org/10.1177/0165025411425873

Kluytmans, A., Van de Schoot, R., Mulder, J., & Hoijtink, H. (2012). Illustrating Bayesian evaluation of informative hypotheses for regression models. *Frontiers in Psychology*, *3*, 1–11. https://doi.org/10.3389/fpsyg.2012.00002

Kwon, O. N., Park, J. H., & Park, J. S. (2006). Cultivating divergent thinking in mathematics through an open-ended approach. *Asia Pacific Education Review*, *7*, 51-61. https://doi.org/10.1007/BF03036784

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174. https://doi.org/10.2307/2529310

Leikin, R. (2009). Exploring mathematical creativity using multiple solution tasks. In R. Leikin, A. Berman, & B. Koichu (Eds.), *Creativity in mathematics and the education of gifted students* (pp. 129-145). Sense Publishers.

Linacre, J. M. (1996). Generalizability theory and many-facet Rasch measurement. In G. Engelhard Jr. & M. Wilson (Eds.), *Objective measurement: Theory into practice* (pp. 85-98). Ablex.

Long, H. (2014). An empirical review of research methodologies and methods in creativity studies (2003–2012). *Creativity Research Journal*, *26*, 427-438. https://doi.org/10.1080/10400419.2014.961781

Long, H., & Pang, W. (2015). Rater effects in creativity assessment: A mixed methods investigation. *Thinking Skills and Creativity*, *15*, 13-25. https://doi.org/10.1016/j.tsc.2014.10.004

Ma, H. H. (2009). The effect size of variables associated with creativity: A meta-analysis. *Creativity Research Journal*, *21*, 30–42. https://doi.org/10.1080/10400410802633400

Main, L. F., Delcourt, M. A., & Treffinger, D. J. (2019). Effects of group training in problemsolving style on future problem-solving performance. T*he Journal of Creative Behavior*, *53*, 274-285. https://doi.org/10.1002/jocb.176

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341. https://doi.org/10.1207/s15328007sem1103_2

Mayer, R.E. (2006). The role of domain knowledge in creative problem-solving. In J.C. Kaufman & J. Baer (Eds.), *Creativity and reason in cognitive development* (pp. 145–158). Cambridge University Press.

McCaffrey, T. (2016). A visual representation to quantitate, diagnose, and improve creativity in insight problem solving. *The Journal of Creative Behavior.* https://doi.org/10.1002/jocb.132

Moonen-van Loon, J. M. W., Overeem, K., Donkers, H. H. L. M., Van der Vleuten, C. P. M., & Driessen, E. W. (2013). Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Advances in Health Sciences Education*, *18*, 1087-1102. https://doi.org/10.1007/s10459-013-9450-z

Mulder, J., & Wagenmakers, E. J. (2016). Introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments. *Journal of Mathematical Psychology*, *72*, 1–5. https://doi.org/10.1016/j.jmp.2016.01.002

Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, *46*, 1–39. https://doi.org/10.18637/jss.v046.i02

Mullet, D. R., Willerson, A., Lamb, K. N., & Kettler, T. (2016). Examining teacher perceptions of creativity: A systematic review of the literature. *Thinking Skills and Creativity*, *21*, 9-30. https://doi.org/10.1016/j.tsc.2016.05.001

Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal*, *4*, 91–122. https://doi.org/10.1080/10400419109534380

Mumford, M. D., Baughman, W. A., Supinski, E. P., & Maher, M. A. (1996a). Process-based measures of creative problem-solving skills: II. Information encoding. *Creativity Research Journal*, *9*, 77–88. https://doi.org/10.1207/s15326934crj0901_6

Mumford, M. D., Baughman, W. A., Threlfall, K. V., Supinski, E. P., & Costanza, D. P. (1996b). Process-based measures of creative problem-solving skills: I. Problem construction. *Creativity Research Journal*, *9*, 63–76. https://doi.org/10.1207/s15326934crj0901_6

Mushquash, C., & O'Connor, B. P. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods*, *38*, 542-547. https://doi.org/10.3758/bf03192810

Muthén, L.K., & Muthén, B.O. (2012). Mplus user's guide. Muthén & Muthén. Myszkowski, N., & Storme, M. (2019). Judge response theory? A call to upgrade our psychometrical account of creativity judgments. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 167-175. https://doi.org/10.1037/aca0000225

Myszkowski, N., Storme, M., Davila, A. and Lubart, T. (2015). Managerial creative problem solving and the big five personality traits. *Journal of Management Development*, *34*, 674-684. https://doi.org/10.1108/JMD-12-2013-0160

Nathoo, F. S., & Masson, M. E. J. (2016). Bayesian alternatives to null-hypothesis significance testing for repeated-measures designs. *Journal of Mathematical Psychology*, *72*, 144–157. https://doi.org/10.1016/j.jmp.2015.03.003

Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2014). Ready, set, create: What instructing people to "be creative" reveals about the meaning and mechanisms of divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 423–432. https://doi.org/10.1037/a0036549

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational Assessment*, *9*, 45–53. https://doi.org/10.1177/073428299100900104

Oral, G., Kaufman, J. C., & Agars, M. D. (2007). Examining creativity in Turkey: Do Western findings apply? *High Ability Studies*, *18*, 235–246. https://doi.org/101080/13598130701709590.

Peele, H. (2006). Appreciative inquiry and creative problem solving in cross-functional teams. *Journal of Applied Behavioral Science*, *42*, 447–467. https://doi.org/10.1177/0021886306292479

Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, 7, 258–264. https://doi.org/10.1016/j.tsc.2012.04.009.

Plucker, J. A., & Beghetto, R. A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 153–167). American Psychological Association. https://doi.org/10.1037/10692-009

Plucker, J., Beghetto, R. A., & Dow, G. (2004). Why isn't creativity more important to educational psychologists? Potential, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*, 83–96. https://doi.org/10.1207/s15326985ep3902_1

Poddiakov, N. (2011). Searching, experimenting and the heuristic structure of a preschool child's experience. *International Journal of Early Years Education*, *19*, 55-63. https://doi.org/10.1080/09669760.2011.571000

Poulou, M. (2001). The role of vignettes in the research of emotional and behavioural difficulties. *Emotional and Behavioural Difficulties*, *6*, 50–62. https://doi.org/10.1080/13632750100507655

Pretz, J. E., & McCollum, V. A. (2014). Self-perceptions of creativity do not always reflect actual creative performance. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 227–236. https://doi.org/10.1037/a0035597

Primi, R., Silvia, P. J., Jauk, E., & Benedek, M. (2019). Applying many-facet Rasch modeling in the assessment of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 176-186. https://doi.org/10.1037/aca0000230

Pringle, A., & Sowden, P. T. (2017). Unearthing the creative thinking process: Fresh insights from a think-aloud study of garden design. *Psychology of Aesthetics, Creativity, and the Arts*, *11*, 344–358. https://doi.org/10.1037/aca0000144

Puccio, G. J., Burnett, C., Acar, S., Yudess, J. A., Holinger, M., & Cabra, J. F. (2018). Creative problem solving in small groups: The effects of creativity training on idea generation, solution creativity, and leadership effectiveness. *The Journal of Creative Behavior*, *54*, 453-471. https://doi.org/10.1002/jocb.381

Qian, M., Plucker, J. A., & Yang, X. (2019). Is creativity domain specific or domain general? Evidence from multilevel explanatory item response theory models. *Thinking Skills and Creativity*, *33*, 100571. https://doi.org/10.1016/j.tsc.2019.100571

Redifer, J. L., Bae, C. L., & Zhao, Q. (2021). Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking. *Learning and Instruction*, *71*, 101395. https://doi.org/10.1016/j.learninstruc.2020.101395

Reiter-Palmon, R. (2017). The role of problem construction in creative production. *The Journal of Creative Behavior*, *51*, 323–326. https://doi.org/10.1002/jocb.202

Reiter-Palmon, R., & Arreola, N. J. (2015). Does generating multiple ideas lead to increased creativity? A comparison of generating one idea vs. many. *Creativity Research Journal*, *27*, 369-374. https://doi.org/10.1080/10400419.2015.1087274

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 144–152. https://doi.org/10.1037/aca0000227

Reiter-Palmon, R., Illies, M. Y., Cross, L. K., Buboltz, C., & Nimps, T. (2009). Creativity and domain specificity: The effect of task type on multiple indexes of creative problem-solving. *Psychology of Aesthetics, Creativity, and the Arts*, *3*, 73–80. https://doi.org/10.1037/a0013410

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2010). The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British Journal of Psychology*, *101*, 47–68. https://doi.org/10.1348/000712609X414204.

Ritter, S. M., Gu, X., Crijns, M., & Biekens, P. (2020). Fostering students' creative thinking skills by means of a one-year creativity training program. *PLoS One*, *15*, e0229773. https://doi.org/10.1371/journal.pone.0229773

Rowe, P. G. (1987). *Design thinking.* MIT Press.

Runco, M. A. (1991). The evaluative, valuative, and divergent thinking of children. *The Journal of Creative Behavior*, *25*, 311–319. https://doi.org/10.1002/j.2162-6057.1991.tb01143.x

Runco, M. A. (2011). *Creativity testing: Manual for the Runco Creativity Assessment Battery.* Creativity Testing Services.

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, *24*, 66–75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., Acar, S., & Cayirdag, N. (2017). A closer look at the creativity gap and why students are less creative at school than outside of school. *Thinking Skills and Creativity*, *24*, 242-249. https://doi.org/10.1016/j.tsc.2017.04.003

Runco, M. A., Illies, J. J., & Eisenman, R. (2005). Creativity, originality, and appropriateness: What do explicit instructions tell us about their relationships? *The Journal of Creative Behavior*, *39*, 137–148. https://doi.org/10.1002/j.2162-6057.2005.tb01255.x

Said-Metwaly, S., Kyndt, E., & Noortgate, W. (2017). Approaches of measuring creativity: A systematic literature review. *Creativity. Theories – Research – Applications*, *4*, 238–275. https://doi.org/10.1515/ctra-2017-0013

Saxon, J. A., Treffinger, D. J., Young, G. C., & Wittig, C. V. (2003). Camp Invention®: A creative, inquiry-based summer enrichment program for elementary students. *The Journal of Creative Behavior*, *37*, 64–74. https://doi.org/10.1002/j.2162-6057.2003.tb00826.x

Schoevers, E. M., Kroesbergen, E. H., & Kattou, M. (2020a). Mathematical creativity: A combination of domain-general creative and domain-specific mathematical skills. *The Journal of Creative Behavior*, *54*, 242-252. https://doi.org/10.1002/jocb.361

Schoevers, E. M., Leseman, P. P., & Kroesbergen, E. H. (2020b). Enriching mathematics education with visual arts: Effects on elementary school students' ability in geometry and visual arts. *International Journal of Science and Mathematics Education*, *18*, 1613-1634. https://doi.org/0.1007/s10763-019-10018-z

Scott, G., Leritz, L. E., & Mumford, M. D. (2004). Types of creativity training: Approaches and their effectiveness. *The Journal of Creative Behavior*, *38*, 149-179. https://doi.org/10.1002/j.2162-6057.2004.tb01238.x

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Sage.

Silvia, P. J. (2008). Discernment and creativity: How well can people identify their most creative ideas? *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 139–146. https://doi.org/10.1037/1931-3896.2.3.139

Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*, 139- 148.

https://doi.org/10.1037/a0014940
Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 68–85. https://doi.org/10.1037/1931-3896.2.2.68
Smith E. V. Jr., & Kulikowich, J. M. (2004). An application of generalizability theory and many-facet Rasch measurement using a complex problem-solving skills assessment. *Educational and Psychological Measurement, 64*, 617-639. https://doi.org/10.1177/0013164404263876
Solomon, P. G. (2009). *The curriculum bridge: From standards to actual classroom practic*e. SAGE.
Sommer, U., Fink, A., & Neubauer, A.C. (2008). Detection of high ability children by teachers and parents: Psychometric qualities of new rating checklists for the assessment of intellectual, creative and social ability. *Psychology Science Quarterly, 50*, 189–205.
Sophonhiranrak, S., Suwannatthachote, P., & Ngudgratoke, S. (2015). Factors affecting creative problem solving in the blended learning environment: A review of the literature. *Procedia-Social and Behavioral Sciences, 174*, 2130-2136. https://doi.org/10.1016/j.sbspro.2015.02.012
Sowden, P. T., Pringle, A., & Peacock, M. (2020). Verbal protocol analysis as a tool to understand the creative process. In V. Dörfler & M. Stierand (Eds.) *Handbook of research methods on creativity* (pp. 314-328). Edward Elgar Publishing.
Sternberg, R. J. (2000). The concept of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 3–15). Cambridge University Press.
Sternberg, R. J. (2012). The assessment of creativity: An investment-based approach. *Creativity Research Journal, 24*, 3–12. https://doi.org/10.1080/10400419.2012.652925
Sternberg, R. J. (2020). What's wrong with creativity testing?. *The Journal of Creative Behavior, 54*, 20-36. https://doi.org/10.1002/ jocb.237
Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences.* Routledge.
Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2004). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing, 9*, 239-261. https://doi.org/10.1016/j.asw.2004.11.001
Sun, M., Wang, M., & Wegerif, R. (2020). Effects of divergent thinking training on students' scientific creativity: The impact of individual creative potential and domain knowledge. *Thinking Skills and Creativity, 37*, 100682. https://doi.org/10.1016/j.tsc.2020.100682
Tomesen, M., Engelen, R., & Hiddink, L. (2019). *Wetenschappelijke verantwoording LVS toetsen Begrijpend Lezen 3.0 voor groep 8* [Scientific justification of the reading comprehension test for Grade 6]. Cito.
Torrance, E. P. (1972). Predictive validity of the Torrance Tests of Creative Thinking. *The Journal of Creative Behavior, 6*, 236–252. https://doi.org/10.1002/j.2162-6057.1972.tb00936.x
Treffinger, D. (1995). Creative problem solving: Overview and educational implications. *Educational Psychology Review, 7*, 301–312. https://doi.org/10.1007/BF02213375
Treffinger, D. J. (2000). *Practice problems for creative problem solving.* Prufrock Press, Inc.
Treffinger, D. J., Selby, E. C., & Isaksen, S. G. (2008). Understanding individual problem-solving style: A key to learning and applying creative problem solving. *Learning and Individual Differences, 18, 390–401*. https://doi.org/10.1016/j.lindif.2007.11.007
Troman, G., & Woods, P. (2001). *Primary Teachers' Stress.* Psychology Press.

Urban, M., & Urban, K. (2022). Orientation toward intrinsic motivation mediates the relationship between metacognition and creativity. *The Journal of Creative Behavior*, *57*, 6-16. https://doi.org/10.1002/jocb.558

Urhahne, D. (2011). Teachers' judgments of elementary students' ability, creativity, and task commitment. *Talent Development & Excellence*, *3*, 229–237.

Van der Vleuten, C. P. M., Dolmans, D. H. J. M., & Scherpbier, A. J. J. A. (2000). The need for evidence in education. *Medical Teacher*, *22*, 246–250.

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 842–860. https://doi.org/10.1111/cdev.12169

Van Harpen, X. Y., & Sriraman, B. (2013). Creativity and mathematical problem posing: an analysis of high school students' mathematical problem posing in China and the USA. *Educational Studies in Mathematics*, *82*, 201–221. https://doi.org/10.1007/s10649-012-9419-5

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modeling cognitive processes.* Academic.

Van Vugt, J., De Vos, T., Milikowski, M., & Milikowski, R. (2019). *Boom LVS Rekenen-Wiskunde.* Boom Uitgevers.

Walton, A. P., & Kemmelmeier, M. (2012). Creativity in its social context: The interplay of organizational norms, situational threat, and gender. *Creativity Research Journal*, *24*, 208-219. https://doi.org/10.1080/10400419.2012.677345

Webb, M. E., Little, D. R., Cropper, S. J., & Roze, K. (2017). The contributions of convergent thinking, divergent thinking, and schizotypy to solving insight and non-insight problems. *Thinking & Reasoning*, *23*, 235-258. https://doi.org/10.1080/13546783.2017.1295105

Willemsen, R.H., Schoevers, E.M. & Kroesbergen, E.H. (2020). The structure of creativity in primary education: An empirical confirmation of the Amusement Park Theory. *Journal of Creative Behavior*, *54*, 857-870. https://doi.org/10.1002/jocb.411

Yekovich, F. R., Thompson, M. A., & Walker, C. H. (1991). Generation and verification of inferences by experts and trained nonexperts. *American Educational Research Journal*, *28*, 189–209. https://doi.org/10.3102/00028312028001189

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal*, *23*, 24-37. https://doi.org/10.1080/10400419.2011.545713

Zhu, Y., Ritter, S. M., & Dijksterhuis, A. (2020). Creativity: Intrapersonal and interpersonal selection of creative ideas. *The Journal of Creative Behavior*, *54*, 626-635. https://doi.org/10.1002/jocb.397

R

CURRICULUM VITAE &
LIST OF PUBLICATIONS AND
PROFESSIONAL CONTRIBUTIONS

“Put an umbrella on the box, to make sure the sun cannot reach it.”

Curriculum Vitae

Mare van Hooijdonk was born in 1988 in Nijmegen. After finishing her vwo in secondary school, she completed her teaching training program for primary education (pabo) at the University of Applied Sciences in Utrecht. During her pabo studies, she volunteered for three months at a boarding school in Nepal and completed half of the pre-master educational sciences at Utrecht University. After completing the last few courses of this pre-master, she decided to pursue her studies with the research master Educational Sciences: Learning in Interaction. This study sparked her interest in educational research. She obtained a PromoDoc funding to combine her passion for teaching and research for 6,5 years. As a teacher, she worked at Basisschool de Verwondering and primarily taught children aged 4 to 7 here. She conducted her PhD studies at Utrecht University, at the department of Education (Faculty of Social Sciences). Her interest in teaching and creativity eventually resulted in this dissertation on the assessment of creative problem solving in primary school students. Since September 2021, she works as a teacher and researcher at Radboud University, Nijmegen. She primarily teaches students from the master's program Educational Sciences and students from the academic teacher training program for primary education (ALPO). Furthermore, she pursues her research within the IMpact of Activies in Gifted Education (IMAGE) project, studying how teachers, parents and care providers can collaborate to help gifted students with complex educational needs. From April 2023 she will continue this work at Radboud University as an Assistant Professor (UD) and also conduct further research on creativity in education.

Peer Reviewed Publications

Van Hooijdonk, M., Ritter, S. M., Linka, M., & Kroesbergen, E. (2022). Creativity and Change of Context: The influence of object-context (in)congruency on cognitive flexibility. *Thinking Skills and Creativity*, *45*, 101044. https://doi.org/10.1016/j.tsc.2022.101044

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2022). Examining the assessment of creativity with generalizability theory: An analysis of creative problem solving assessment tasks. *Thinking Skills and Creativity*, *43*, 100994. https://doi.org/10.1016/j.tsc.2021.100994

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2020). Creative problem solving in primary education: Exploring the role of fact finding, problem finding, and solution finding across tasks. *Thinking Skills and Creativity*, *37*, 100665. https://doi.org/10.1016/j.tsc.2020.100665

Kroesbergen, E.H., Van Hooijdonk, M., van Viersen, S., Middel-Lalleman, M., & Reijnders, J. (2016). The psychological well-being of early identified young gifted children. *Gifted Child Quarterly*, *60*, 16-30. https://doi.org/10.1177/0016986215609113

Conference Presentations and Proceedings

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2022) *Assessing Creative Problem Solving in Primary School Students.* UPCE seminar (online)

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2022) The Alignment between Teacher *Assessments and Trained Rater Assessments of Primary School Students' Creative Problem Solving Abilities.* PSN Conference (online)

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2022) *Assessing Creative Problem Solving in Primary Education.* MIC Conference, Bologna

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2019) *Creatief probleem-oplossen Meetbaar Maken in het Basisonderwijs.* ORD Heerlen

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2018) *A Generalizability Study with Creative Problem Solving Data*, Creativity from an Embodied and Situated View, Utrecht.

Van Hooijdonk, M., Kroesbergen, E. H., & Van Tartwijk, J. (2016) *Creative Problem Solving: Exploring the role of fact finding, problem finding, and solution finding across tasks.* Empowering Creativity in Education Conference, Utrecht.

Van Hooijdonk, M., Kroesbergen, E. H., & Van Tartwijk, J (2016) *Exploring Creative Problem Solving in Primary School Students in Two Domains.* MIC Conference, Bologna

Professional Contributions

Van Hooijdonk, M., Verduijn, N., & Mainhard, M., (2021). Vakkennis én creativiteit: een leerbaar duo. *Didactief Online.*

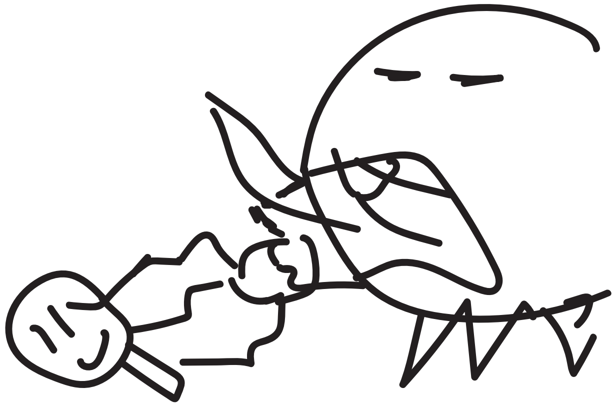
Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2020) *Creatief probleem-oplossen Meetbaar Maken in het Basisonderwijs.* Workshop PO Conferentie Utrecht

Van Hooijdonk, M., Kroesbergen, E. H., & Van Tartwijk, J. (2017) *Creativiteit in het Basisonderwijs.* Workshop POINT Utrecht

Oomen, C., Van Hooijdonk, M., & Smit, K. (2014). *De Lerende Organisatie. Wat is het en hoe geef je er vorm aan?* Universiteit Utrecht.

Prast, E. J., van Hooijdonk, M., van de Weijer-Bergsma, E., Kroesbergen, E. H. & Van Luit, J.E.H. (2013). Wat vinden kinderen van hun rekenlessen? Rekenen met plezier. *JSW: Jeugd in School en Wereld*, *98*, 6-9.

DANKWOORD



“Ask a pokemon to freeze-up the ice pops.”

Dit proefschrift gaat over het creatief oplossen van problemen. Binnen mijn promotietraject ben ik, logischerwijs, ook tegen diverse problemen aangelopen. Gelukkig heb ik deze problemen met een groep fantastische mensen, al dan niet creatief, kunnen aanpakken. Hieronder zou ik graag nog een aantal van deze mensen en de hulp die ze geboden hebben willen noemen.

Jan, welk probleem los jij eigenlijk niet op? Een promotietraject combineren met een baan voor de klas? Jan regelt PromoDoc. Een artikel heeft proofreading nodig? Dochter Fransje vliegt in vanuit Cambridge (waarvoor ontzettend veel dank!). Geen vervoer van de kroeg naar het station? Spring maar bij je promotor achterop. Je hebt me daarnaast geleerd dat veel problemen voorkomen worden door te zorgen dat je met leuke mensen werkt. Nou Jan, je bent er zelf zo een.

Evelyn, bij jou is er altijd ruimte voor problemen. En het mooie is, soms hoeven ze niet eens direct te worden opgelost. Ze mogen er gewoon zijn. Ik heb alle ruimte ervaren om zelf te zoeken naar oplossingen en waar nodig jouw hulp in te roepen. Daarnaast heb ik natuurlijk enorm genoten van de MIC conferenties die we samen in Bologna bezochten. Zouden ze in Griekenland ook wijn bij de lunch serveren?

Tim. Voordat ik jou kende konden mijn schrijfskills nog wel een update gebruiken. Na jouw guerilla-de-eerste-10-regels tactiek ging het stukken beter. Eigenlijk was dit zo met al mijn academische vaardigheden. Als ik dacht dat ik mijn grens wel bereikt had, dan wist jij deze altijd nog iets op te rekken en me nog meer te leren. Daarnaast wil ik je graag nog bedanken voor het regelen van de geweldige vader van mijn kinderen.

De keuze voor mijn paranimfen was niet moeilijk. Eén persoon beschouw ik als de ultimate-problem-solver: David. Ga een dag met David op zijn werkkamer zitten en je merkt meteen dat hij niet alleen jou, maar ook de rest van de afdeling van allerhande oplossingen voorziet. Daarnaast ben je ook nog eens een fijn persoon, die met een dosis humor en eerlijkheid durft te zeggen waar het op staat. Er is niemand waarmee ik liever friet eet.

Als er iemand is die zichzelf - structureel – op de tweede plek zet, dan ben jij het wel, Noor. Ik geloof niet dat ik iemand met een groter hart voor mensen én voor onderwijs ken. Jij ziet in iedere leerling die creatieve vonk. Je hebt dit promotietraject van begin tot eind van dichtbij meegemaakt en bij iedere stap weer je enorme vertrouwen in mij uitgesproken. Als dit promotietraject me iets heeft opgeleverd, dan is het wel een vriendin voor het leven. Ik ben blij dat je vandaag naast me staat.

Jan Willem, toen PromoDoc een optie werd, kwam ik met mijn 'probleem' bij jou: ik zocht een fijne school waarbij ik het lesgeven en het doen van onderzoek kon combineren. Al snel mocht ik met jou om de tafel. Ons ruim anderhalf uur durende sollicitatiegesprek ging vijftien minuten over mij en mijn plan. De rest van ons gesprek ging over onderwijsontwikkeling en over mooie plannen voor de toekomst. Ik was geloof ik aangenomen. De afgelopen jaren hebben we deze inhoudelijk gesprekken aangevuld met persoonlijke gesprekken en ik hoop dat we dit nog jaren blijven doen.

Mijn lieve collega's van Basisschool de Verwondering. Ik heb eigenlijk nooit problemen ervaren bij het combineren van mijn onderzoek met mijn werk voor de klas. Ja, een enkel probleem misschien: dat ik graag overal bij wilde zijn op school. Wat zijn jullie een warm bad. Dat je als schoolteam echt kan (en moet!) samenwerken, dat laten jullie keer op keer zien. Ik mis jullie nog iedere dag en ben blij dat ik nog regelmatig maffe prentenboeken en liedjes van Sandra Kim met jullie kan delen. Het was daarnaast heel fijn dat er zoveel ruimte voor mijn onderzoek was op school. Ik voelde me altijd welkom op Leerplein 3, om materialen uit te proberen en om daadwerkelijk data te verzamelen. Inez, Aziza en Maartje, daarvoor wil ik jullie nog even extra bedanken.

Waar veel collega's de nodige creatieve oplossingen moeten bedenken om voldoende participanten te kunnen werven voor hun onderzoek, was dat voor mij eigenlijk niet zo'n probleem. Mede dankzij de steun van Conexus en de vele fijne contacten in het basisonderwijs waren er snel voldoende scholen bereid deel te nemen. Bedankt leerkrachten en leerlingen voor jullie deelname!

Dan is er nog het probleem van 'te weinig tijd en te weinig handen'. Als ik alle klassen zelf had moeten bezoeken, dan was ik denk ik nu nog bezig geweest. Suus, bedankt voor al je hulp en je flexibiliteit. Daarnaast moest er enorm veel data voor mijn onderzoek dubbel gescoord worden. Joris en Isabelle, bedankt voor al jullie uren assistentie!

UU Educatie. Wat heb ik een goede tijd gehad bij jullie. F3.01 was een heerlijke plek om steeds weer te kunnen landen. Alle problemen kregen hier alle ruimte. We hebben samen kunnen sparren, lachen, een traan gelaten en life events kunnen delen zoals bruiloften en geboortes (zoals die van Krakje). Handig ook dat we in de laatste maanden alle regelzaken rondom de promotie met elkaar konden delen. Dit geldt ook voor onze club met PromoDocs. Mooi dat we 'peers' hadden in ons combinatietraject en veel van elkaars ervaringen hebben kunnen leren. Gelukkig is de wereld van de onderwijswetenschappen niet al te groot en zullen we elkaar nog regelmatig treffen.

Creativiteit. Daar ligt mijn onderzoekshart. Hoe leuk is het om iedere keer meer te leren over iets waarbij oplossingen per definitie niet vast staan. Tijdens mijn promotie heb ik deze passie met een club mede-onderzoekers kunnen delen. 'The creativity group': Honghong, Eveline, Marloes, Marije en later Isabelle, Robin en Kim. UPCE in the Utrechtse of United vorm, Labroup of Creativity in Education, CreaNea… we hebben wat namen gehad. We hebben veel gelezen en vanalles georganiseerd. We combineerden gezelligheid met inhoud. Samen met jullie heb ik veel mogen leren. En daar gaan we mee door. Ik kijk uit naar onze sessie op de Earli en ben blij dat ik met enkelen van jullie, Ard en Evelyn ons creativiteitsonderzoek voort kan zetten en zo nog veel meer mag leren.

Werken met leuke mensen voorkomt dus problemen, of zorgt er in ieder geval voor dat ze minder groot lijken. Mijn lieve collega's van onderwijswetenschappen en PW aan de RU, het is fantastisch om jullie (weer!) te leren kennen en om met jullie onderwijs te mogen verzorgen. Goed dat we elkaar scherp houden, waar nodig steunen en iedere dag ons onderwijs een beetje beter proberen te maken. En natuurlijk heel prettig dat er ook tijd is voor een persoonlijk praatje of grapje binnen en buiten het partyhok. Daarnaast wil ik hierbij nog mijn 'gifted' collega's van de RU noemen. Fijn dat we binnen IMAGE en TWS samen kunnen werken en dat ik altijd bij jullie binnen kan stormen, met én zonder problemen.

Een groot deel van dit proefschrift is in de coronatijd geschreven. We weten allemaal dat dit voor veel mensen een eenzame periode is geweest. Ik heb het intense geluk gehad dat ik niet veel van deze sociale problemen heb ervaren door een club hele fijne mensen om me heen. Een bakje koffie in de voortuin, een burenborrel op de stoep, een lekker toetje voor de voordeur. En zodra het weer kon een nachtje borrelen en logeren bij een hotel, inclusief inhaal-Kerstfeest. Ik ben blij dat we nu onze reguliere buurtborrels en weekendjes weg weer voort kunnen zetten. En Marie, hoe geweldig is het dat jij als buurvrouw mijn proefschrift hebt kunnen vormgeven. Dat maakt die uren opmaken achter de computer toch stukken gezelliger.

Pap en mam, jullie hebben mij op de middelbare school regelmatig geplaagd met mijn motivatie-probleem. Het heeft jullie wel eens verbaasd dat iemand die zó weinig deed voor school, toch zoveel van onderwijs is gaan houden. Ik wil jullie, maar ook Fenna, Daan, Thomas en mijn hele schoonfamilie bedanken voor het feit dat problemen binnen onze familie altijd gedeeld kunnen worden en dat iedereen bereid is om mee te denken om ze op te lossen. En dat jullie altijd een keertje extra willen oppassen als ik weer eens wat moet voor werk.

Gijs. Zonder jou was mijn leven waarschijnlijk een opstapeling van problemen geweest. Dan zou ik veel te vaak achter de computer zitten, me altijd mee laten slepen in het moment, overal te laat zijn en geen schone broek meer in mijn kast hebben liggen. Jij bent de motor bij ons thuis, door jou blijft alles draaien. Los van dat zorg jij met jouw nuchterheid en humor dat ogenschijnlijk grote problemen ineens niet meer zo groot en juist heel hanteerbaar lijken. We houden allebei van ons werk, van onze kinderen, van elkaar en van al onze bezigheden daarbuiten. Ik denk dat we het allemaal heel goed organiseren samen.

Lien en Tuur. Jullie zijn misschien wel het mooiste en belangrijkste resultaat van mijn promotie-onderzoek. Mijn grootste probleem is misschien wel dat ik niet kan werken én - tegelijkertijd - bij jullie kan zijn. Maar gelukkig zijn jullie heerlijk flexibel en genieten jullie minstens net zoveel van de dagen bij papa, de opvang en opa en oma. Lien, ik weet dat je liever had dat ik een boek schreef over 'twee vechtende eekhoorntjes', maar we moeten het toch maar hier mee doen. Dat zou overigens wel een veel creatievere verdediging worden. Misschien iets voor de toekomst?

Dan komen we nu bij mijn laatste probleem. Ik kan simpelweg niet alle mensen noemen die me tijdens dit traject gevormd en gesteund hebben. Daarom nog een enorm woord van dank voor alle andere mensen die ik op dit onderzoekspad heb getroffen. Ook heel veel dank voor de beoordelingscommissie die deze laatste stap, de verdediging, mogelijk maken. Met veel plezier heb ik de afgelopen jaren een dit proefschrift gewerkt. Het afsluiten van deze periode is voor mij dan ook bitterzoet. Zoet omdat de afsluiting van dit proefschrift weer nieuwe deuren opent. Bitter omdat ik het intensieve samenwerken met mijn promotieteam en de mensen eromheen zal missen. Maar we zoeken wel een creatieve manier om samen te kunnen blijven werken, toch?

D