# Examining Kindergarten Children's Testing and Optimising in the Context of a Gear Engineering Task

Timo Reuter [1]*, Miriam Leuchter [1]

[1] *University of Koblenz-Landau*, GERMANY

***Corresponding Author:** reutertimo@uni-landau.de

## ABSTRACT

Introducing kindergarten children to the engineering design process (EDP) is an important objective of early STEM education. Studies indicate that children often miss the crucial steps of testing and optimising during the EDP and do not persist in making solutions better. The present study explores how children's goal awareness, self-evaluation ability, domain-specific content knowledge, spatial skills and intelligence relate to their persistence, testing and optimising behaviour and to solution quality. In a standardized procedure, 41 children (4 to 7 years) in Germany worked on an engineering task in the domain of gears. The engineering process was videotaped, children's testing and optimising, solution quality, goal awareness, self-evaluation, and task persistence were rated with a coding scheme and with interviewer questions. Domain-specific content knowledge, mental rotation ability and figural reasoning were measured with standardized tests. Correlational analyses indicated that goal awareness was positively related to solution quality. However, most children required support by the interviewer to retrieve the goal specifications. Moreover, children's self-evaluation was negatively related to task persistence. Most children were satisfied with their first solution, even when it did not meet the requirements. Our findings emphasize the important role of teachers in helping children to tackle challenges with the EDP.

**Keywords:** individual differences, problem-solving, kindergarten, gears, engineering

## INTRODUCTION

Scientists and practitioners agree that STEM (science, technology, engineering, and mathematics) education should begin as early as kindergarten (e.g., NRC, 2013; OECD, 2018) as it can provide children with a fundamental knowledge base for learning STEM subjects in later school years (Kaderavek et al., 2020). Early STEM education aims at promoting children's understanding of domain-specific content knowledge, fostering domain-general process skills, and establishing an understanding of epistemic practices (NRC, 2013). In recent years, engineering has been increasingly emphasized as a central component of early STEM education (Bustamante et al., 2018; John et al., 2018). For instance, the *Next Generation Science Standards* framework in the US integrated engineering into science education by raising engineering design to the same level as scientific inquiry, and thus emphasises how closely the two are linked to each other (NRC, 2013; on the close link see also Schauble et al., 1991). For instance, an analysis and interpretation of data is required in order to (i) understand relations between causes and effects (scientific inquiry) and (ii) to determine if a designed product works as intended to solve a problem (engineering design). Moreover, there is a growing body of research exploring young children's engineering skills in both formal learning settings, such as preschool and elementary school (Bairaktarova et al., 2011; Gold et al., 2020; Kendall,

2015; Lottero-Perdue and Tomayko, 2020; Strimel et al., 2018a), and informal learning settings, such as museum exhibits (Eshan et al., 2021; Marcus et al., 2021; Tõugu et al., 2017).

In engineering education, particular emphasis is placed on introducing children to the engineering design process (EDP) as the disciplinary core idea of engineering (NRC, 2013). The EDP has evolved as the systematic and iterative problem-solving strategy engineers use to identify and solve problems (Cross, 2008; Lucas et al., 2014; Pleasants and Olson, 2019). The EDP is systematic in that it follows a characteristic sequence of steps comprising (i) identifying a problem, (ii) planning solutions, (iii) building models and prototypes, and (iv) evaluating solutions and prototypes (Cross, 2008). The EDP is iterative in that engineers build various prototypes, continuously test and optimise their solutions, analyse and interpret data, learn from failures, and make decisions based on evidence in order to improve solution quality (Cross, 2008; Lucas et al., 2014). Design failures, i.e. solutions that fail to meet one or more criteria, and the need to revise designs are regarded normal parts of the process (Petroski, 1992). Thus, persisting and learning from failure is considered as a crucial epistemic practice of engineering (Crismond and Adams, 2012; Lucas et al., 2014). Studies comparing the design processes of experts and novices reveal that expert designers go through significantly more iterations than novices, resulting in increased quality of the engineering solution (Atman et al., 2007; Strimel et al., 2018b). However, there are only a few empirical studies that have examined the EDP with focus on the crucial practices of persisting, testing and optimising in kindergarten and elementary school-aged children. Learning more about young children's persistence, testing and optimising, is highly relevant because of its theoretical implications and because it might also inform practitioners for developing targeted intervention programs based on these findings (Strimel et al., 2018a). In the present exploratory study, we therefore investigate kindergarten children's task persistence, testing and optimising when solving an engineering design task.

## Empirical Studies Examining Children's Task Persistence, Testing and Optimising

Testing and optimising require the ability to analyse and interpret data (NRC, 2013), which develops between the ages of 3 and 7 (Zimmerman and Klahr, 2018). Children as young as 3 years are capable of spontaneously producing evidence that would enable causal learning (Schulz et al., 2007). However, studies observing children's engineering indicate that pre-schoolers and kindergartners do not sufficiently analyse and interpret data since they tend to neglect testing and optimising (for an overview, see Gold et al., 2020). For instance, Bairaktarova et al. (2011) stress that 3- to 5-year-olds' testing during free play activities with unstructured materials, such as sandboxes or water tables, semi-structured materials, such as paints and paper, and structured play materials, such as snap circuits focused mainly on the completion of a single construction, rather than on evaluating different prototypes to improve the solution quality. Similarly, in a study by Strimel et al. (2018a) 5- to 7-year-old kindergarten children spent only a very small part of their effort on evaluating solutions when solving engineering design tasks. Moreover, most children ended their design session once a single prototype was built (Strimel et al., 2018a). So far, little is known about young children's persistence, and neglection of testing and optimising. However, the few studies indicate that kindergarteners' persistence, testing and optimising might be related to the problem type (well- vs. ill-defined), children's goal awareness, ability to realistically self-evaluate solutions, and domain-specific content knowledge (Kendall, 2015; Lottero-Perdue and Tomayko, 2020; Strimel et al., 2018a). In the following, we will briefly address these points.

Children might need to be aware of the goal in order to test and optimise. Strimel et al. (2018a) suggest that the low incidence of persisting, testing and optimising in their study might have been due to the fact that they provided children with ill-defined engineering tasks. In ill-defined engineering tasks, the goal and constraints are not given in advance; thus, the criteria for evaluating a solution must be determined by the problem solver (Dörner and Funke, 2017). As kindergarten children often find it difficult to identify such criteria, well-defined engineering tasks seem more appropriate to investigate kindergarten children's testing and optimising (Haluschak et al., 2018; Strimel et al., 2018a). In well-defined engineering tasks, there is a clearly defined goal that allows for multiple solutions, and the criteria to evaluate a solution are given in advance (Crismond, 2001). In a museum's activity context, Pagano et al. (2019) investigated how the setting of a goal affected family's conversation about engineering practices. Families in a program with a set goal reflected more often on testing, things not working, optimising, and being successful or unsuccessful than families who completed an activity without a set goal.

Nevertheless, well-defined engineering tasks require the problem solver to build and maintain a mental representation of the goal state, which places high demands on the problem solvers' working memory (Kintsch, 2007). As the capacity of working memory in kindergarten children is still developing (Diamond, 2013), children might have problems in building and maintaining adequate mental representations of a goal state. Karmiloff-Smith (1979) demonstrated that 4- to 9-year-olds built accurate initial mental representations of a problem's goal state, but tended to simplify the problem by modifying their mental representation of the goal state during their problem-solving process (Karmiloff-Smith, 1979). In a qualitative study with a well-defined engineering task, Kendall (2015) investigated how six kindergarten children evaluated two design solutions of bridges. When asked to recall the

requirements the bridge had to meet (length and sturdiness), most children did recall only one of the two requirements. Consequently, studies on kindergarten children's persisting, testing and optimising should control whether children are aware of a task's goal and the task's requirements.

Learning from failures by continuously testing and optimising prototypes requires the ability to assess one's own performance and solutions realistically. Findings from developmental psychology indicate that self-evaluations of 4- to 7-year-olds tend to be unrealistically positive (Harter and Bukowski, 2015; Marsh et al., 2012; Oppermann et al., 2018). Such a bias in perception might lead children not to optimise, as there is nothing to optimise from their point of view. When the children in the study by Kendall (2015) were asked to test and optimise the engineered bridges, children tended to focus on what worked, but neglected failures and their sources. Contrary, in a study by Lottero-Perdue and Tomayko (2020) 5- to 7-year-old kindergarten children were able to accurately self-evaluate their engineered solution with respect to the requirements it had to meet, and almost all of the children opted to optimise their design when asked by the researchers. As the authors acknowledge, the engineering task was quite straightforward with respect to success or failure: The children were asked to build a fence out of wooden and foam blocks to contain a small, fast-, and randomly-moving *HEXBUG nano*® robot and give the robot as much room as possible to move. The children could hardly ignore design failures as in that case the robot evidently broke through the fence. Most children were able to detect failures' causes (e.g., the foam blocks were too light), however, an accurate failure analysis did not automatically lead the children to apply this knowledge in a revised design solution (Lottero-Perdue and Tomayko, 2020). Thus, research on kindergarten children's engineering behaviour might investigate how children's self-evaluation of their solutions relate to their persistence, testing and optimising.

Professional engineers apply scientific and mathematical content knowledge to design, test and optimise solutions (Pleasants and Olson, 2019). For young children, however, the relationship between domain-specific content knowledge and testing, optimising and engineering outcome is less straightforward. On the one hand, studies with 3- to 5-year-olds suggest that children are able to carry out meaningful tests and manipulations on a novel toy, use this evidence to infer causal structures, and thus, learn how the toy works without prior domain-specific knowledge (Schulz et al., 2007). Thus, engaging children in engineering practices such as testing and optimising might provide learning opportunities for relevant domain-specific content knowledge (Lin et al., 2020; Penner et al., 1997).

On the other hand, museum studies with 4- to 9-year-old children and their parents in two conditions (brief instruction on relevant domain-specific content knowledge or not) showed that instructed families displayed more engineering-related talk, and their children made use of the domain-specific knowledge more often than in the other condition (Benjamin et al., 2010; Marcus et al., 2021). To the best of our knowledge, engineering studies with young children so far did not measure children's prior-knowledge in the relevant science domain. In the present study, we therefore measure prior-knowledge and examine how it relates to task persistence, testing, optimising and solution quality.

Moreover, spatial ability has been found to be strongly associated with the development of expertise in STEM over the course of education (Wai et al., 2009). With respect to engineering, spatial skills have been found to be closely linked to physical object manipulation (such as testing and making changes to constructions) for 9- to 13-year-old children (Ramey and Uttal, 2017). The correlation between spatial skills and engineering is likely to be greater the more the engineering task requires the problem-solver to mentally manipulate information about objects in the environment and the motion of objects (Tõugu et al., 2017). In the present study, the children were given a task in which the rotation of gears had to be considered. Consequently, mental rotation ability could possibly play a role and was therefore assessed.

Furthermore, fluid intelligence is one of the most important prerequisites for learning and closely interrelated with reasoning and problem-solving (Diamond, 2013). To the best of our knowledge, studies on engineering in kindergarten did not control for children's fluid intelligence and how it might relate to persistence, testing, optimising and solution quality.

**Present Study**

In the present exploratory study, we aim at learning more about how kindergarten children's persistence, testing and optimising in an engineering task is related to children's goal awareness, their self-evaluation of the solution, and their domain-specific content knowledge, spatial skills, and fluid intelligence. As previous studies mostly used more qualitative methods providing fine-granulated insights, we would like to complement these qualitative analyses with a more quantitative approach by measuring and statistically correlating children's persistence, testing and optimising, goal awareness, self-evaluation of the solution, and the objective solution quality to learn more about the interrelations. Additionally, we measure children's content knowledge in the relevant domain, and include mental rotation ability as an indicator of spatial skills as well as figural reasoning as an indicator of fluid intelligence in our analysis. We address the following research questions (RQs):
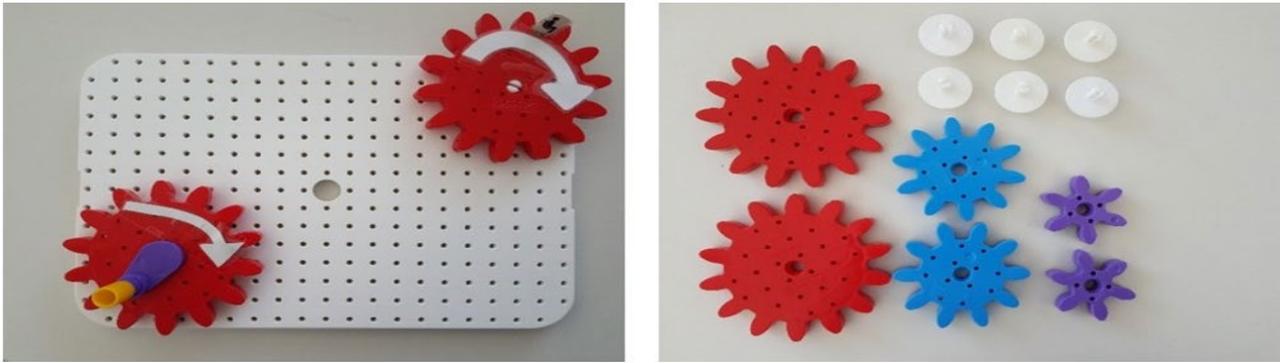
**Figure 1.** Peg board driving gear and "carousel" | construction materials

1. **RQ-1:** What solution quality do the children achieve? (solution quality)
2. **RQ-2:** Do the children test and optimise? (testing and optimising)
3. **RQ-3:** Are the children aware of the goal? (goal awareness)
4. **RQ-4:** Are the children satisfied with their solution? (self-evaluation)
5. **RQ-5:** Do the children make full use of the given time for engineering? (time-on-task as an indicator of task persistence)
6. **RQ-6:** Do the children want to make a change to their solution when explicitly asked? (willingness to make changes as an indicator of task persistence)
7. **RQ-7:** How are solution quality, testing, optimising, goal awareness, self-evaluation, task persistence, domain-specific content knowledge, spatial skills, and fluid intelligence related?

We expect that the better the children's goal awareness is, the more they will test and optimise, which will correspond with a higher time-on-task, and result in better solution quality. Moreover, we assume a negative correlation between children's self-evaluation and their willingness to make changes to their solution.

## METHODS

### Participants

Forty-one children (18 girls and 23 boys) participated in the study. We recruited the children by convenience sampling in seven kindergartens in a rural part of Germany. All participants were informed about the goals of the study. The children participated voluntarily and with the written consent of their parents. The convenience sampling yielded the following age distribution: Three children were 4 years old (one child was 51 months, two children were 58 months), 17 children were 5 years old, another 17 children were 6 years old, and two children were 7 years old (82 months and 91 months). The age information is missing for two children. Given the explorative nature of our study, we deliberately decided to allow the wide age range to get as much variance as possible in our data.

### Engineering Task

We developed a well-defined engineering design task in the domain of gears. Gears seem to be a suitable domain to study 4- to 7-year-old children's testing and optimising for at least three reasons: (i) Children can easily construct, test and modify gears with age-appropriate (toy) gear-kits. (ii) Gears allow for formulating tasks with clear goals (well-defined problems), e.g., making a target gear spin in a given direction with specified turning direction of a driving gear. Moreover, the children can directly and constantly test the achievement of such a goal by turning the gears and receive immediate feedback by observing the resulting turning direction. (iii) Gears are an elementary component of mechanics and incorporate domain-specific content knowledge. Developmental studies suggest that children differ with respect to their content knowledge about gears' turning direction (Reuter and Leuchter, 2021; Lehrer and Schauble, 1998). Among 4- to 7-year-olds, there are children who know that meshed gears turn in opposite directions, children who have naïve concepts (meshed gears turning in the same direction), and children who have no apparent concepts (Reuter and Leuchter, 2021).

In a one-to-one setting, the children received a plastic peg board (31cm×22cm) with two red plastic gears (approximately 12 cm in diameter with 14 teeth) that were permanently fixed on the board. One gear was on the left-hand side at the front of the peg board and had a crank (driving gear). An arrow pointing to the right was glued on the driving gear. The other gear was on the right-hand side further back on the peg board. This gear also had an arrow on it pointing to the right. Additionally, a sitting stick figure made of paper looking in the direction of the arrow was placed at the gear (see **Figure 1**).

The interviewer told the children that this gear was a 'carousel' with a person on it. Then, the interviewer instructed the children that the 'carousel' should turn forward, i.e., in the direction of the arrow (specification 1) when the driving gear is turned in the direction of the arrow (specification 2). As constructional materials, the children received two red gears that were the same size as the driving gear and carousel (large gear), two medium-sized blue gears (approximately 9cm in diameter with ten teeth), two small purple gears (approximately 6 cm in diameter with six teeth) and six plugs (see **Figure 1**). The peg board had prefabricated holes that allowed one to easily mount the gears with a plug. The interviewer pointed out to the children that she or he could use all of the materials, but they did not have to do so. It was possible to connect the driving gear with the carousel using two gears (e.g., the two large red ones). The intended difficulty for the children was to realize that this straightforward (and maybe the most obvious) solution would result in the wrong turning direction of the carousel; thus, it would not meet specification 1. The task can be characterized as a well-defined engineering task since it had a clear goal state, i.e., clearly defined specifications (turning direction of the carousel with a given turning direction of the driving gear); it demanded goal-oriented thinking under constraints (the driving gear and carousel were fixed, and the construction materials and space were restricted); and it allowed for more than one possible solution (e.g., connecting driving gear and carousel with three or with five gears, and there were different possible arrangements of the gears).

### Test Procedure

The engineering design task was conducted in one-to-one situations in a quiet room in the children's kindergartens. The interviewer and the child were sitting side by side at a table or on the floor. It was carried out using a standardized procedure, consisting of three or four steps: (i) introduction, (ii) engineering trial 1, (iii) assessment of task persistence, goal awareness and probing for children's self-evaluation of their solution and, if applicable (iv) engineering trial 2. **Table 1** describes the procedure of the engineering task in detail. The complete procedure was videotaped.

**Table 1.** Procedure of the engineering task

| | Step | Interviewer does … | Interviewer says … |
|---|---|---|---|
| Introduction | 1 | Places an empty peg board, two yellow large size gears, two blue medium-sized gears, two purple small size gears, and plugs on the table. | Today, you can play with gears. Here, you have various gears. Now you get to test how to put the gears on the peg board. |
| | 2 | Removes the mounted gears from the peg board. | Okay, now you know how it works! |
| | 3 | Places the peg board with the fixed driving gear and fixed "carousel" on the table. | |
| | 4 | Points to the driving gear. | Look, here is a gear with a crank. You can turn it. |
| | 5 | Points at the "carousel" gear. | The gear here is a carousel. You see, there is a person sitting on it. |
| | 6 | (i) turns the carousel in the direction of the arrow with his finger, and (ii) turns the driving gear in the direction of the arrow with his finger | *Introducing the well-defined problem:* Now, your task is this: (i) The carousel should rotate in the direction of the arrow so that the person on the carousel moves forward (ii) when you turn the crank in the direction of the arrow. Turn the crank in the direction of the arrow! That's right, you can turn this way. Then, the person should move forward on the carousel. Turn the carousel forward! Okay, that's how the carousel should go. You can use all the gears and plugs for building - but you do not have to. Do you understand the task? Do you have any questions? Since more children are supposed to get their turn with this game today, I have to look at the clock. When the time is up, I will tell you and we will look at what you have built. If you have a product before then, you let me know. Okay, start now! |
| Trial 1 | 7 | Stops the time (max. 3 minutes) | |

**Table 1 (Continued).**

| | Step | Interviewer does … | Interviewer says … |
|---|---|---|---|
| **Assessment** | 8 | | *Exploring task persistence:* [If child finishes before the three minutes are up]: Okay, you have a product you want to show me yet? I'll look in a minute.<br>[If the three minutes are up]: Okay, I'm afraid the time is up. I can't wait to see what you've built! |
| | 9 | | *Probing goal awareness:* Before we look at your construction, tell me again: What was the task?<br>[If the child did not provide a complete description of the task, the interviewer probed for the missing specifications]: E.g., What direction should the carousel turn? |
| | 10 | | *Probing self-evaluation:* Are you satisfied with your solution? |
| | 11 | | *Probing task persistence:* Do you want to change something?<br>[If yes]: Okay, then you can build again! |
| **Trial 1** | 12 | Stops the time (max. 2 minutes) | |
| | 13 | | Okay, you did a good job! Now we're done for today. |

**Table 2.** Scoring solution quality

| Score | | Description |
|---|---|---|
| 3 points | Complete, adequate prototype | Driving gear and carousel are connected with an *odd* number of gears |
| 2 points | Complete, inadequate prototype | Driving gear and carousel are connected with an *even* number of gears |
| 1 point | Incomplete prototype | Gears have been built on the driving gear and/or the carousel, but driving gear and carousel are not connected |
| 0 points | No prototype | Neither driving gear nor carousel are meshed with another gear |

Note. When a child builds a complete prototype but then makes an alteration making it incomplete again at the 3 minutes mark of trial 1 (or 2 minutes mark of trial 2), it would be coded as an incomplete prototype.

### Coding Scheme for the Engineering Design Task

To analyse the children's testing and optimising, we developed a coding scheme in a two-step process. First, all categories of the coding scheme were derived in a top-down approach. We therefore adapted parts of the *Metacognitive Skills in Constructional Play Engagement (MetaSCoPE)* coding scheme (Spektor-Levy et al., 2017). Second, two raters independently tested the coding scheme with 20 videos. To obtain interrater agreement, we calculated Cohen's kappa (Cohen, 1960). In an iterative process, the coding scheme was revised in a bottom-up procedure, uncertainties were discussed by the raters and decision criteria were clarified. The final coding scheme comprised a code to rate the solution quality, four dichotomous codes indicating children's testing and optimising, and two codes to record the children's goal awareness.

#### Solution quality

We rated the quality of the prototype the children had built with a score ranging from 0 to 3 points (see **Table 2**) after trial 1 and, if applicable, after trail 2. Rater agreement was substantial with Cohen's kappa=.767. However, given the clear coding scheme, this appears to be a rather low agreement. In some cases, it was not clear from the videos whether two (or more) gears were meshing or not. We decided for these cases to always code as if the gears were meshing (i.e., to give the higher score).

#### Testing and optimising

We coded indications of testing and optimising for trial 1 and, if applicable, for trial 2. To account for the findings of Bairaktarova et al. (2011) and Strimel et al. (2018a), we distinguished between children's testing of incomplete prototypes (e.g., turning two meshed gears) and of complete prototypes (turning the driving gear to move the carousel) during the trial. Following the same logic, we distinguished between changes children made to incomplete and to complete prototypes during the trial. Cohen's kappa was .614; thus, according to Landis and Koch (1977), rater agreement can be regarded as substantial.

**Table 3.** Correlations

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | solution quality[a] | | | | | | | | | |
| 2 | testing[b] | **.713** | | | | | | | | |
| 3 | optimising[b] | **.459** | **.681** | | | | | | | |
| 4 | goal awareness[c] | **.316** | .140 | **.444** | | | | | | |
| 5 | self-evaluation[b] | .180 | -.167 | -.356 | -.007 | | | | | |
| 6 | task persistence (time-on-task)[b] | -.009 | **.453** | **.479** | -.078 | **-.573** | | | | |
| 7 | task persistence (willingness to make changes)[b] | -.218 | .032 | .229 | -.037 | **-.677** | .004 | | | |
| 8 | content knowledge gears' turning direction[a] | .034 | -.056 | .039 | .290 | -.362 | -.099 | .168 | | |
| 9 | mental rotation ability[a] | **.336** | .238 | .233 | .103 | -.227 | -.221 | .186 | -.069 | |
| 10 | figural reasoning ability[c] | .131 | -.056 | .210 | .015 | .033 | -.168 | .267 | -.005 | **.374** |

Note. [a]Interval, normally distributed; [b]Cntinuous dichotomy; [c]Interval, non-normally distributed; Significant correlations (p<.05) are highlighted in bold.

### *Goal awareness*

We recorded and coded whether the children were able to name specification 1 (the carousel should turn in the direction of the arrow) without the help of the interviewer, with help or not at all. We did the same for specification 2 (the driving gear has to be turned in the direction of the arrow). The final rater agreement was substantial with Cohen's kappa=.754. We calculated a score to quantify children's goal awareness. For each of the two specifications, the children received 2 points if they provided the correct answer on their own, 1 point if they provided the correct answer after the probing of the interviewer, and 0 points if they did not give a correct answer even after probing. Thus, the score ranged from 0 to 4 points.

### Testing Domain-Specific Knowledge, Spatial Skills and Figural Reasoning

We tested children's domain-specific conceptual knowledge of gears' turning direction in a standardized one-to-one interview (gear-interview). The children were shown eight different gear configurations (items) built with physical gears. Each item consisted of a driving gear with an arrow indicating the turning direction that was meshed with one or more other gears (target gears). The children had to predict the turning direction of the target gears. All gears were fixed, that is, they could not rotate. This meant that the children could not learn during the test. Moreover, there was no teaching on gear functioning before. For a detailed description of the procedure see Authors. The internal consistency was high with $a_{KD20}$=.912. The children could achieve a maximum score of 26 points.

We assessed mental rotation ability with the *Picture Rotation Test (PRT)* (Quaiser-Pohl, 2003) as an indicator for spatial skills. Children were presented with coloured three-dimensional target images of people (e.g., an ice-skater) and animals (e.g., a tiger) as well as three comparison images. Two of them were mirrored and rotated and one was only rotated. The children had to identify the rotated image of the target image. The maximum score was 16 points.

As an indicator of fluid intelligence, we measured figural reasoning with the matrices subtest from the *Culture Fair Test (CFT)* (Weiß and Osterland, 2013). The children could reach a maximum score of 15 points. All tests were conducted about one week before the engineering task to ensure adequate testing time and cognitive loads for the age group.

### Correlational Analysis

For the correlational analysis, we considered the different measurement levels of the variables and whether the data were distributed normally or not (c.f. Field, 2011; see Table 3). We calculated Pearson's correlation coefficient $r$ for pairs of variables that were normally distributed at the interval level. We calculated Spearman's correlation coefficient $\varrho$ for pairs of variables with one or both variables non-normally distributed at the interval level. We calculated the biserial correlation coefficient $r_b$ for pairs of variables where one of the variables was measured dichotomously but had an underlying continuum, and the other variable was normally distributed at the interval level. In cases of non-normally distributed data of the interval level variable, we calculated the biserial correlation coefficient based on Spearman's correlation coefficient $\varrho$. We calculated the tetrachoric correlation coefficient $r_{tet}$ for pairs of dichotomously measured continuous variables.

# RESULTS

## Mean Values and Proportions

### *RQ 1: Solution quality*

The mean quality of children's solutions in trial 1 was 1.61 (SD=1.02, range=0-3). Nine children built a complete and adequate prototype, 14 children a complete but inadequate prototype. Eleven children had an incomplete prototype at the end of trial 1 (max. 3 minutes), and seven children had no prototype. For the subsample of the children who did trial 2 (n=23), the mean score slightly increased from 1.48 (SD=0.95) after trial 1 to 1.70 (SD=0.88) after trial 2. However, the mean scores of trial 1 and trial 2 did not differ significantly, t(22)=-1.417, p=.171, as a paired sample t-test revealed. Thus, only trial 1 will be considered in the further analyses.

### *RQ 2: Testing and optimising*

We first analysed children's testing and optimising regardless of the completeness of the prototype. Thirty children tested at least once. Twenty-six children made at least one change. To examine whether children's testing behaviour differed for incomplete and complete prototypes, we restricted our analysis to children who finished trial 1 with a complete prototype (n=23). Sixteen of them tested the functioning of both incomplete and complete prototypes. Four of them tested the functioning of complete prototypes only, and one child the functioning of incomplete prototypes only. Two of them did not test at all. To examine whether children's optimising behaviour differed for incomplete and complete prototypes, we restricted our analysis to children who finished trial 1 with a complete, but inadequate prototype (n=14). Six of them made changes to both incomplete and complete prototypes, another six children made at least one change to an incomplete, but not to a complete prototype, and two of them did not make any changes.

### *RQ 3: Goal awareness*

The mean score for goal awareness was 2.24 (SD=1.07, range=0-4). Five children (12%) provided a complete and correct description of both specifications on their own. 27 children (66%) provided an incomplete description of the task and did not mention the turning direction. However, with probing of the interviewer, they were able to give the complete description. Thus, there were a total of 32 children (78%) who could correctly name both specifications. The remaining nine children (22%) provided incomplete or no description of one or of both specifications, despite the interviewer's probing.

### *RQ 4: Self-evaluation*

When asked by the interviewer, 31 children (75%) said that they were satisfied with their solution in trial 1.

### *RQ 5 and RQ 6: Task persistence*

Sixteen children (39%) made full use of the time allowed at trial 1 and thus were interrupted by the interviewer after 3 minutes of engineering. Correspondingly, the other 25 children (61%) presented a solution to the interviewer before the 3 minutes were over. Their mean time-on-task was 114 seconds (SD=41.8). When asked by the interviewer, 23 children (56%) opted to make changes to their solution. Hence, these 23 children did trial 2. Nine of these children made full use of the two minutes that were allowed in trial 2. The other 14 children finished before the two minutes were over with a mean time-on-task of 76 seconds (SD=32.8). The mean time-on-task for the children who did trial 1 only was 134 seconds (SD=51.2), and it was 237 seconds (SD=58.4) for the children who did both trial 1 and trial 2.

## Correlational Analyses

### *RQ 7: correlations*

The solution quality was positively related to testing, to optimising, to goal awareness and to children's mental rotation ability. Testing was positively related to optimising, and to time-on-task. Optimising was positively related to goal awareness and to time-on-task. The children's self-evaluation of their solution was negatively related to time-on-task. Moreover, self-evaluation was negatively related to the children's willingness to make a change to their solution. Mental rotation ability (M=9.78, SD=4.11) was positively correlated with figural reasoning ability (M=4.29, SD=2.72). Domain-specific content knowledge on gears' turning direction (M=11.49, SD=5.64) did not significantly correlate with any of the other variables. Table 3 gives an overview of the correlation matrix.

**Table 4.** Association between solution quality and testing and optimizing, goal awareness, self-evaluation, task persistence, and content knowledge

| Solution quality after trial 1 | n | ... tested at least once | ... optimised at least once | ... were aware of the goal | | ... were satisfied with the solution (self-evaluation) | ... made full use of the 3 minutes (task persistence) | ... wanted to make changes when asked after trial 1 (task persistence) | Content knowledge gears' turning direction | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Total¹ | without probing | | | | ... had a correct concept of gears' turning direction | ... had an incorrect concept of gears' turning direction | ... had no apparent concept of gears' turning direction |
| | n | (n=30) | (n=26) | (n=32) | (n=5) | (n=31) | (n=16) | (n=23) | (n=9) | (n=14) | (n=18) |
| 3 pts. | 9 | 9 | 6 | 8 | (3) | 9 | 2 | 4 | 3 | 3 | 3 |
| 2 pts. | 14 | 12 | 12 | 12 | (2) | 9 | 7 | 6 | 4 | 6 | 4 |
| 1 pt. | 11 | 7 | 7 | 7 | (0) | 7 | 6 | 10 | 2 | 3 | 6 |
| 0 pt. | 7 | 2 | 1 | 5 | (0) | 6 | 1 | 3 | 0 | 2 | 5 |

Note. 3 pts.: Complete, adequate prototype; 2 pts.: Complete, inadequate prototype; 1 pt.: Incomplete prototype; 0 pt.: No prototype; 1 number of children who named both specifications with or without probing of interviewer.

Our next analysis is concerned with the association between solution quality and all other variables (**Table 4**). However, with respect to content knowledge, the mean value does not adequately represent the concepts. Thus, we categorize the children as either having the correct concept (meshed gears turn in opposite directions), an incorrect concept (meshed gears turn in the same direction), or no apparent concept by using the binomial distribution (P26;1/2 [X=17]=8.43%, meaning that 17 out of 26 consistent responses have 8.43% probability of occurring by chance, i.e., by guessing). Children with at least 17 out of 26 correct predictions in the gear interview were considered as having the correct concept (n=9). Children with at least 17 out of 26 incorrect predictions were considered as having the incorrect concept (n=14). Children without a consistent answer pattern were considered as guessing, thus having no apparent concept (n=18).

Due to the small sample size, Fisher-Freeman-Halton's exact test was used to determine if there was a significant association between solution quality and the respective variable (Bortz and Lienert, 2008). In the case of a significant association, we performed post-hoc pairwise comparisons with Bonferroni corrected significance level of p<.008. There was a statistically significant association between solution quality and testing, $\chi^2$=10.942, p=.006. The higher the solution quality, the higher the proportion of children who tested at least once. Pairwise comparisons revealed that children with a complete and adequate prototype (3 points) differed significantly (Fisher's exact test, p=.005) from children with no prototype (0 points) in their testing behaviour (9/9 vs. 2/7). Furthermore, there was a statistically significant association between solution quality and optimising, $\chi^2$=9.776, p=.016. Pairwise comparisons revealed that children with a complete, but inadequate prototype (2 points) differed significantly (Fisher's exact test, p=.003) from children with no prototype (0 points) in their optimising behaviour (12/14 vs. 1/7). There was no statistically significant association between solution quality and any of the other variables (all p>.05).

However, on a descriptive level, there was a U-shaped association between solution quality and self-evaluation: For both 3 points and 0 points, the proportion of children who were satisfied with their constructions was high (9/9 and 6/7), whereas for both 2 points and 1 point the proportion was comparably low (9/14 and 7/11). Likewise, on a descriptive level, an inverted U-shaped association could be observed between solution quality and task persistence: For both 3 points and 0 points, the proportion of children making full use of the three minutes was rather low (2/9 and 1/7), whereas for both 2 points and 1 point the proportion was comparably high (7/14 and 6/11). Furthermore, 10 out of 11 children with 1 point in solution quality wanted to make a change to their construction when asked by the interviewer after trial 1. For all other children, this proportion was lower with 43 %.

With respect to children's content knowledge of gears' turning direction, descriptive data show on the one hand: The better the solution quality, the greater the proportion of children with the correct concept. On the other hand, the data also show: Among the children with an adequate prototype (3 points), 3 out of 9 had the correct concept of gears' turning direction. Another 3 achieved the 3 points with an incorrect concept, and the remaining 3 children without an apparent concept.

# DISCUSSION

In our exploratory study, we attempted to learn more about kindergarten children's persistence, testing and optimising in an engineering task and how these behaviours are related to children's goal awareness, their self-evaluation of the solution, their domain-specific content knowledge, spatial skills, and fluid intelligence. Therefore, we had 4- to 7-year-old children work on a well-defined engineering task with more than one possible correct solution in the domain of gears. We rated the solution quality (RQ-1); observed the children's engineering behaviour for incidences of testing and optimising (RQ-2); and measured goal awareness (RQ-3), self-evaluation of the solution (RQ-4), and task persistence (RQ-5 and RQ-6). Moreover, we included the children's domain-specific content knowledge, mental rotation ability and figural reasoning ability and explored how all aspects were related to each other (RQ-7). The discussion section aims at integrating the results of the research questions and attempts to draw a coherent picture of the findings.

## Testing, Optimising and Task Persistence

The majority of the children showed at least one indication of testing and optimising in the course of the engineering trial. As expected, both testing and optimising did correspond with a higher time-on-task and with increased solution quality. These findings are in line with studies showing that experts achieve better solutions than novices by going through more design iterations (Atman et al., 2007; Strimel et al., 2018b). Thirty-nine percent of the children made full use of the 3 minutes. Other than expected, there was no significant linear correlation between time-on-task and solution quality. Descriptive data showed that children who succeeded to build an adequate prototype did not necessarily make full use of the time. This indicates that the task was easy enough for these children. Moreover, children who did not succeed in the task at all did not make full use of the time as well. This might be an indicator that the task was not sufficiently interesting for these children (Crismond, 2001) or that they did not succeed in inhibiting other stimuli (Miyake et al., 2000). Approximately 50% of the children opted to make changes to their solution when explicitly asked by the interviewer. There was no significant correlation between willingness to make changes and solution quality. Moreover, as descriptive data showed, the proportion of children who wanted to make a change to their construction was highest among those with an incomplete prototype, whereas the proportion was low for children with a complete, but inadequate prototype. These findings suggest that children's testing and optimising mainly aims at the completion of one single construction rather than at constructing and evaluating different prototypes, which is in line with other studies (Bairaktarova et al., 2011; Strimel et al., 2018a). Interestingly, a substantial proportion of the children (43%) who built a correct prototype wanted to continue building and making changes. We cannot say from our data whether this reflects what Lucas et al. (2014) calls the engineering habit of mind of relentlessly trying to make things work better, or if the children just enjoyed building with the gears without pursuing such a goal.

We examined various aspects that might be related to children's persistence and their testing and optimising behaviour. We will discuss them in the following sections.

### *Goal awareness*

As shown in a study by Kendall (2015), children had difficulties in recalling the two specifications on their own. Indeed, the majority of the children stated independently from the interviewer that the carousel should rotate when they turn the crank, but they specified the turning direction only when explicitly asked for. This suggests that when children were introduced to the task, they built an accurate mental representation of the goal state (Kintsch, 2007), but they might have 'simplified' this goal representation during the course of the 3-minute engineering phase (Karmiloff-Smith, 1979). Even after the probing of the interviewer, one out of five children did not provide a complete description of how the construction was supposed to work. These children may have built an incomplete initial mental model of the goal state, or may have difficulties in maintaining an adequate representation of the goal even with well-defined problems. This difficulty may be related to the limited capacity of working memory in 4- to 7-year-old children (Diamond, 2013). Future studies should therefore examine the role of working memory and other executive functions for engineering in early childhood, e.g., using the *EF touch battery* (Willoughby et al., 2013). As expected, goal awareness was positively related to solution quality and optimising. Other than expected, we found no significant correlation between goal awareness and testing. This lack of correlation can be interpreted as children having tested regardless of a correct and complete goal representation.

### *Self-evaluation*

Overall, the majority of the children were satisfied with their solution. Moreover, as expected, the children who were satisfied with their solution were less willing to make changes to their prototype when asked by the interviewer. However, there was no significant linear correlation between self-evaluation and solution quality.

Descriptive data rather showed indications of a u-shaped relation: both the children who built an adequate prototype and the children who finished trial 1 with constructions showing no evidence of a prototype were satisfied. In contrast, the proportion of satisfied children was lower among those who built an incomplete prototype or a complete but inadequate prototype. These findings are partly in line with research showing that self-evaluations of 4- to 7-year-olds might be unrealistically positive (Harter and Bukowski, 2015; Marsh et al., 2012; Oppermann et al., 2018). This bias in self-evaluation could be one reason why kindergarten children use little or no iterative approach when working on an engineering design task (Gold et al., 2020; Strimel et al., 2018a). However, as indicated by findings of Kendall (2015) and Lottero-Perdue and Tomayko (2020), children are willing to improve their solutions if the requirements are made clear to them and if they are instructed to test their solutions. In this respect, a methodological shortcoming of our study is the interviewer's question *"Are you satisfied with your solution?"*, which might be too general. If the question was put in relation to the demanded task specifications (e.g., *"Does your carousel turn in the demanded direction?"* or *"Compare your solution to the demanded requirements."*), the children might be able to evaluate their solution in depth. However, since we stated the requirements a few seconds before, when probing for children's goal awareness, we assume that it was possible for the children to relate our question to the task specifications. Nevertheless, a future study could examine whether children come to a more realistic assessment if they are given an explicit criterion for evaluation (Crismond, 2001; Lottero-Perdue and Tomayko, 2020).

### Domain-specific content knowledge

We found no significant correlations of domain-specific content knowledge with the other variables in our data. This result indicates that it was possible for the children to test and optimise and to succeed in building an adequate prototype, even without specific knowledge of meshed gears' turning directions. As can be seen in the descriptive data, among the children with the correct solution, one third of them had a correct, an incorrect or no apparent concept of gears' turning direction respectively. However, as Lottero-Perdue and Tomayko (2020) also note in their study, task characteristics might have elicited this result. In our study, it obviously was possible to solve the task in a short time. Exploring the gears and observing their behaviour thoroughly might have been sufficient to succeed. An indication of the relevance of this interpretation can be found in Schulz et al.'s (2007) findings that children are able to produce evidence to infer causal structures of a mechanism by thorough testing. However, the descriptive data also show that the better the solution, the greater the proportion of children with the correct concept on gears' turning direction. This is in line with studies indicating that prior-knowledge in the relevant science domain can contribute to better engineering processes and outcomes (Benjamin et al., 2010; Marcus et al., 2021). Future studies should therefore try to differentiate whether successful solutions are achieved through trial and error procedures or through the application of content knowledge. Moreover, upcoming research might investigate and compare kindergarten children's engineering behaviour in different science domains, e.g., the stability of block buildings (Weber et al., 2020), one-sided levers as simple machines (Leuchter and Naber, 2018), floating of objects (Hong and Diamond, 2012), and pursue two directions: (1) Studies should examine for different science domains whether kindergarten children can learn domain-specific content knowledge by being exposed to engineering design tasks (Lin et al., 2020; Penner et al., 1997). (2) Studies might compare children who received a training in the relevant domain-specific knowledge beforehand with children who did not receive a training (Benjamin et al., 2010; Marcus et al., 2021). This will provide insights into children's failure analysis and the role of domain-specific content knowledge for optimising.

### Spatial skills and fluid intelligence

Mental rotation ability did positively correlate to figural reasoning ability. Moreover, mental rotation ability was positively related to solution quality, indicating that children with better spatial skills achieved better solutions. This is in line with research suggesting spatial skills to be positively related with success in STEM (Wai et al., 2009). However, our study showed no significant correlation between spatial skills and testing and optimising. This is in contrast to findings from observations with older children (Ramey and Uttal, 2017). Furthermore, figural reasoning was not significantly related to any of the other variables. These findings indicate that children can engage with the engineering design process from an early age regardless of their spatial skills and their fluid intelligence, which is in line with observational studies of young children's engineering behaviour (for an overview, see Gold et al., 2020).

Our results highlight some practical implications. With respect to engineering intervention programs for kindergarten children, our findings emphasize the important role of teachers. They should learn techniques that help children to tackle challenges they might face with respect to persisting, testing and optimising, such as asking the children whether they remember the goal, helping them to retrieve the goal specifications, stating that nobody succeeds in the first attempt, and motivating them to remain on track (Lottero-Perdue and Tomayko, 2020).

**Limitations**

First, in the present study, the test power was limited; thus, we might not have discovered relations in our sample that exist in the population. A larger sample size would also allow us to apply path models to describe the directed dependencies among the variables. Moreover, with a larger sample size it might be worthwhile to examine whether there are (latent) "engineering profiles" among young children, e.g., children that stand out by their talent for thinking up creative solutions, whereas others are better at a critical evaluation, testing and optimisation of a solution. Second, we used only one engineering task in one domain (gears) with one aspect (turning direction). Future studies could use problems on the relative turning speed of gears. Third, we interpreted children's goal awareness, self-evaluation, and task persistence from our observations of the problem-solving process and from the children's answers to the interviewer's questions. This procedure seems to be appropriate as a first step in order to obtain an impression of which aspects might be relevant for children's testing and optimising behaviour. However, future studies should use standardized and validated test instruments to assess children's executive functioning, self-evaluation, and general task persistence. Fourth, the procedure we used in the present study did not allow us to measure competencies developed or demonstrated in group settings, which are important both in school settings and engineering practice, where communication and teamwork are essential (Lucas et al., 2014). Future studies might therefore use group settings as well.

However, to date, not many studies have been conducted in this area and with this age group. Therefore, we assume that the knowledge gained from our exploratory study is worthwhile. We have given the children a well-defined problem, conducted and videotaped the problem-solving process in a standardized way, developed a coding scheme to analyse children's testing and optimising and examined control variables. Thus, we were able to relate persistence and testing and optimising to solution quality and to the control variables. With our study, we have contributed to the development of further studies to investigate young children's EDP competencies as a relevant part of early STEM learning.

## ACKNOWLEDGEMENTS

## REFERENCES

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S. and Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *Journal of Engineering Education*, 96(4), 359-379. https://doi.org/10.1002/j.2168-9830.2007.tb00945.x

Bairaktarova, D., Evangelou, D., Bagiati, A. and Brophy, S. P. (2011). Early engineering in young children's exploratory play with tangible materials. *Children, Youth and Environments*, 21(2), 212-235.

Benjamin, N., Haden, C. A. and Wilkerson, E. (2010). Enhancing building, conversation, and learning through caregiver-child interactions in a children's museum. *Developmental Psychology,* 46(2), 502-515. https://doi.org/10.1037/a0017822

Bortz, J. and Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* [Condensed statistics for clinical research: Guide to nonparametric analysis of small samples]. Berlin, Heidelberg: Springer Berlin Heidelberg.

Bustamante, A., Greenfield, D. and Nayfeld, I. (2018). Early childhood science and engineering: Engaging platforms for fostering domain-general learning skills. *Education Sciences*, 8(3), 144. https://doi.org/10.3390/educsci8030144

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. https://doi.org/10.1177/001316446002000104

Crismond, D. P. (2001). Learning and using science ideas when doing investigate-and-redesign tasks: A study of naive, novice, and expert designers doing constrained and scaffolded design work. *Journal of Research in Science Teaching*, 38(7), 791-820. https://doi.org/10.1002/tea.1032

Crismond, D. P. and Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, 101(4), 738-797. https://doi.org/10.1002/j.2168-9830.2012.tb01127.x

Cross, N. (2008). *Engineering design methods: Strategies for product design*. New York, NY: Wiley. http://www.loc.gov/catdir/enhancements/fy0810/2008002727-d.html

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Dörner, D. and Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8, 1153. https://doi.org/10.3389/fpsyg.2017.01153

Ehsan, H., Rehmat, A. P. and Cardella, M. E. (2021). Computational thinking embedded in engineering design: Capturing computational thinking of children in an informal engineering design activity. *International Journal of Technology and Design Education,* 31(3), 441-464. https://doi.org/10.1007/s10798-020-09562-5

Field, A. (2011). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll).* SAGE.

Gold, Z. S., Elicker, J., Kellerman, A. M., Christ, S., Mishra, A. A. and Howe, N. (2020). Engineering play, mathematics, and spatial skills in children with and without disabilities. *Early Education and Development*, 38(2), 1-17. https://doi.org/10.1080/10409289.2019.1709382

Haluschak, E., Stevens, M., Moore, T., Tank, K., Cardella, M., Hynes, M., Gajdzik, E. and Lopez-Parra, R. (2018). Initial problem scoping in K-2 classrooms (fundamental) [Paper presentation]. *2018 ASEE Annual Conference & Exposition.* Salt Lake City, UT. https://doi.org/10.18260/1-2--30663

Harter, S. and Bukowski, W. M. (2015). *The construction of the self: Developmental and sociocultural foundations.* Guilford Press.

Hong, S.‑Y. and Diamond, K. E. (2012). Two approaches to teaching young children science concepts, vocabulary, and scientific problem-solving skills. *Early Childhood Research Quarterly*, 27(2), 295-305. https://doi.org/10.1016/j.ecresq.2011.09.006

John, M.‑S., Sibuma, B., Wunnava, S., Anggoro, F. and Dubosarsky, M. (2018). An iterative participatory approach to developing an early childhood problem-based STEM curriculum. *European Journal of STEM Education*, 3(3). https://doi.org/10.20897/ejsteme/3867

Kaderavek, J. N., Paprzycki, P., Czerniak, C. M., Hapgood, S., Mentzer, G., Molitor, S. and Mendenhall, R. (2020). Longitudinal impact of early childhood science instruction on 5th grade science achievement. *International Journal of Science Education*, 42(7), 1124-1143. https://doi.org/10.1080/09500693.2020.1749908

Karmiloff-Smith, A. (1979). Problem-solving procedures and children's construction and representations of closed railway circuits. *Archives of Psychology,* XLVII, 37-59.

Kendall, A. (2015). Evaluate-and-redesign tasks: Using interviews to investigate how elementary students iterate (fundamentals), in *122nd ASEE Annual Conference & Exposition*, Seattle, Washington. https://doi.org/10.18260/p.24025

Kintsch, W. (2007). *Comprehension: A paradigm for cognition.* Cambridge University Press.

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

Lehrer, R. and Schauble, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching*, 35(1), 3-25. https://doi.org/10.1002/(SICI)1098-2736(199801)35:1<3::AID-TEA2>3.0.CO;2-X

Leuchter, M. and Naber, B. (2018). Studying children's knowledge base of one-sided levers as force amplifiers. *Journal of Research in Science Teaching*, 33(10), 1099. https://doi.org/10.1002/tea.21470

Lin, X., Yang, W., Wu, L., Zhu, L., Wu, D. and Li, H. (2020). Using an inquiry-based science and engineering program to promote science knowledge, problem-solving skills and approaches to learning in preschool children. *Early Education and Development*, 32(5), 1-19. https://doi.org/10.1080/10409289.2020.1795333

Lottero-Perdue, P. S. and Tomayko, M. (2020). Kindergartners' engagement in an epistemic practice of engineering: Persisting and learning from failure [Paper presentation]. *The ASEE's Virtual Conference 2020.* https://doi.org/10.18260/1-2--34892

Lucas, B., Hanson, J. and Claxton, G. (2014). *Thinking like an engineer: Implications for the education system.* Royal Academy of Engineering.

Marcus, M., Tõugu, P., Haden, C. A. and Uttal, D. H. (2021). Advancing opportunities for children's informal STEM learning transfer through parent-child narrative reflection. *Child Development,* 92(5), e1075-e1084. https://doi.org/10.1111/cdev.13641

Marsh, H. W., Xu, M. and Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham and T. C. Urdan (eds.), *APA educational psychology handbook* (pp. 427-458). American Psychological Association. https://doi.org/10.1037/13273-015

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

NRC. (2013). *Next generation science standards.* National Academies Press. https://doi.org/10.17226/18290

OECD. (2018). *Early learning matters: The international early learning and child well-being study.* https://www.oecd.org/education/school/Early-Learning-Matters-Project-Brochure.pdf

Oppermann, E., Brunner, M., Eccles, J. S. and Anders, Y. (2018). Uncovering young children's motivational beliefs about learning science. *Journal of Research in Science Teaching*, 55(3), 399-421. https://doi.org/10.1002/tea.21424

Pagano, L. C., Haden, C. A., Uttal, D. H. and Cohen, T. (2019). Conversational reflections about tinkering experiences in a children's museum. *Science Education,* 103(6), 1493-1512. https://doi.org/10.1002/sce.21536

Penner, D. E., Giles, N. D., Lehrer, R. and Schauble, L. (1997). Building functional models: Designing an elbow. *Journal of Research in Science Teaching,* 34(2), 125-143. https://doi.org/10.1002/(SICI)1098-2736(199702)34:2<125::AID-TEA3>3.0.CO;2-V

Petroski, H. (1992). *To engineer is human: the role of failure in successful design.* Vintage.

Pleasants, J. and Olson, J. K. (2019). What is engineering? Elaborating the nature of engineering for K-12 education. *Science Education*, 103(1), 145-166. https://doi.org/10.1002/sce.21483

Quaiser-Pohl, C. (2003). The mental cutting test "Schnitte" and the picture rotation test-two new measures to assess spatial ability. *International Journal of Testing*, 3(3), 219-231. https://doi.org/10.1207/S15327574IJT0303_2

Ramey, K. E. and Uttal, D. H. (2017). Making sense of space: Distributed spatial sensemaking in a middle school summer engineering camp. *Journal of the Learning Sciences,* 26(2), 277-319. https://doi.org/10.1080/10508406.2016.1277226

Reuter, T. and Leuchter, M. (2021). Children's concepts of gears and their promotion through play. *Journal of Research in Science Teaching*, 58(1), 69–94. https://doi.org/10.1002/tea.21647

Schauble, L., Klopfer, L. E. and Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching,* 28(9), 859-882. https://doi.org/10.1002/tea.3660280910

Schulz, L. E., Gopnik, A. and Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science,* 10(3), 322-332. https://doi.org/10.1111/j.1467-7687.2007.00587.x

Spektor-Levy, O., Basilio, M., Zachariou, A. and Whitebread, D. (2017). Young children's spontaneous manifestation of self-regulation and metacognition. *Teachers College Record*, 119(13). https://doi.org/10.1177/016146811711901314

Strimel, G. J., Bartholomew, S. R., Kim, E. and Zhang, L. (2018). An investigation of engineering design cognition and achievement in primary school. *Journal for STEM Education Research*, 87(2), 121. https://doi.org/10.1007/s41979-018-0008-0

Strimel, G. J., Kim, E., Bartholomew, S. R. and Cantu, D. V. (2018). Examining engineering design cognition with respect to student experience and performance. *International Journal of Engineering Education*, 34(6), 1910-1929.

Tõugu, P., Marcus, M., Haden, C. A. and Uttal, D. H. (2017). Connecting play experiences and engineering learning in a children's museum. *Journal of Applied Developmental Psychology,* 53, 10-19. https://doi.org/10.1016/j.appdev.2017.09.001

Wai, J., Lubinski, D. and Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817-835. https://doi.org/10.1037/a0016127

Weber, A. M., Reuter, T. and Leuchter, M. (2020). The impact of a construction play on 5- to 6-year-old children's reasoning about stability. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2020.01737

Weiß, R. H. and Osterland, J. (2013). *CFT 1-R.* Hogrefe.

Willoughby, M. T., Pek, J. and Blair, C. B. (2013). Measuring executive function in early childhood: A focus on maximal reliability and the derivation of short forms. *Psychological Assessment*, 25(2), 664-670. https://doi.org/10.1037/a0031747

Zimmerman, C. and Klahr, D. (2018). *Development of scientific thinking.* In J. T. Wixted and S. Ghetti (eds), *Stevens' handbook of experimental psychology and cognitive neuroscience, developmental and social psychology* (pp. 223-247). Newark: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn407