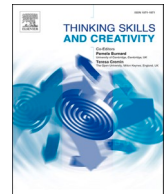




ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Thinking Skills and Creativity

journal homepage: [www.elsevier.com/locate/tsc](http://www.elsevier.com/locate/tsc)

## PROSPER: A comprehensive, valid, and reliable instrument to observe problem-solving behaviours in preschoolers

Joris Van Elsen<sup>a,\*</sup>, Lisanne Buckers<sup>b</sup>, Charlotte Van Tricht<sup>a</sup>, Joke Torbeyns<sup>b</sup>, Sven De Maeyer<sup>a</sup>

<sup>a</sup> University of Antwerp, Faculty of Social Sciences, Department of Training and Education Sciences, Sint-Jacobstraat 2, 2000 Antwerp, Belgium

<sup>b</sup> Faculty of Psychology and Educational Sciences, Centre for Instructional Psychology and Technology, Dekenstraat 2, Postbox 3773, 3000 Leuven, Belgium

### ARTICLE INFO

#### Keywords:

Early childhood education  
Problem solving  
STEM  
Engineering design  
Observation  
Assessment instrument  
Psychometric properties

### ABSTRACT

Problem solving is an important yet challenging competence that is required in STEM education and is highly valued in STEM careers. Learning to solve problems should start at an early age. However, to optimally support young children in developing their problem-solving competencies, there is a need for instruments to measure these competencies in an authentic context. The present study aims to develop the Problem-solving Observation Scheme for Playful Engineering and the Real-world (PROSPER), a comprehensive, valid, and reliable observation scheme to make problem solving visible in preschool children. Twelve preschool children aged 4-6-years individually solved three engineering design problems. All sessions were videotaped, resulting in 17.33 hours of recordings. Five researchers and five preschool teachers identified 1150 problem-solving behaviours through video-based observations. Identified behaviours were coded iteratively, based on the PISA problem-solving framework. Psychometric properties of the observation scheme were analysed. PROSPER consists of 33 behaviours, covering the whole problem-solving process. The PISA framework was extended with four generic, overarching behaviours: commitment, perseverance, autonomy, and responsiveness. Strong intercoder reliability was established. Across task consistency was high with small differences between the three problems. PROSPER proved to be a comprehensive, valid, and reliable instrument to capture preschoolers' problem-solving behaviours in an authentic setting. Future studies can rely on PROSPER to investigate children's development of problem-solving competencies and the effectiveness of classroom interventions.

### 1. Introduction

Problem solving is an important and challenging set of competencies that is required for lifelong learning (European Commission, 2019) and highly valued STEM (Science, Technology, Engineering, and Mathematics) careers (Tytler, 2020). Developing problem-solving competencies takes years. Therefore, fostering problem-solving competencies best starts from early on (e.g., Malcok & Ceylan, 2022; Rogers, 2004). Young children can learn how to solve problems with tailored adult support (Lile Diamond, 2018;

\* Corresponding author Joris Van Elsen, University of Antwerp, Sint-Jacobstraat 2, 2000 Antwerp, Belgium

E-mail addresses: [joris.vanelen@uantwerpen.be](mailto:joris.vanelen@uantwerpen.be) (J. Van Elsen), [lisanne.buckers@student.kuleuven.be](mailto:lisanne.buckers@student.kuleuven.be) (L. Buckers), [charlotte.vantricht@uantwerpen.be](mailto:charlotte.vantricht@uantwerpen.be) (C. Van Tricht), [joke.torbeyns@kuleuven.be](mailto:joke.torbeyns@kuleuven.be) (J. Torbeyns), [sven.demaeyer@uantwerpen.be](mailto:sven.demaeyer@uantwerpen.be) (S. De Maeyer).

<https://doi.org/10.1016/j.tsc.2025.101940>

Received 15 January 2025; Received in revised form 16 July 2025; Accepted 29 July 2025

Available online 30 July 2025

1871-1871/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

Vygotsky, 1978). However, much is still unclear about how problem-solving competencies develop during the preschool years and what interventions are effective in supporting this development, especially when it concerns authentic, real-world problems.

To study the development of problem-solving competencies and the effectiveness of educational interventions, researchers need comprehensive, valid and reliable measurement instruments to assess these competencies. Likewise, educators can benefit from such instruments to systematically monitor the development of preschool children's problem-solving competencies and to tailor their classroom activities to the children's needs.

Since Herbert Simon's (1973) article *The Structure of ill-structured problems*, there is an ongoing debate about the transferability of findings derived from solving well-structured problems (e.g., Tower of Hanoi) to ill-structured problems (e.g., authentic STEM-problems). Findings resulting from research based on well-structured problems may be foundational for studying ill-structured problems (Reed, 2016), but both types of problems require different cognitive processes (Jonassen, 2010). Nowadays, researchers recognise the importance of integrating the context in which the problem is embedded (Csapó & Funke, 2017). Nevertheless, problem-solving competencies are typically assessed using well-structured problems in controlled environments, often removed from the authentic settings of classrooms or homes (Rhodes et al., 2024). There remains a need for instruments that enable researchers and educators to assess real-world problem-solving competencies in context.

Systematic observation of hands-on problem-solving assignments enables the documentation of children's behaviours in naturalistic contexts, capturing both verbal and non-verbal actions without relying too much on children's language skills (Azevedo, 2009). Researchers emphasise the importance of incorporating purposeful observation of children within familiar environments as an integral component of any preschool assessment (Brassard & Boehm, 2007; Monteiro & Jiménez-Aleixandre, 2016).

Engineering design problems are particularly well-suited for observing children's problem-solving behaviours. Engineering design problems are authentic, real-world problems that require the application of STEM ideas to be solved (Arık & Topçu, 2022; Moore et al., 2014). They bear strong similarities to the everyday activities of children as they try to build structures with building blocks and outdoor materials (Geist, 2025; Shechter et al., 2021; Weiser, 2022) or craft things (Kangas et al., 2022).

To the best of our knowledge, there are currently no validated instruments that enable the systematic collection of evidence on preschoolers' problem-solving behaviours, that encompass the full problem-solving process, that can be applied across different problems, and that can be used repeatedly to monitor the development of preschoolers' problem-solving competencies. Therefore, the present study aims to make problem solving visible by developing a comprehensive, valid, and reliable observation instrument that researchers and educators can use to collect evidence on preschoolers' problem-solving competencies in the classroom.

### 1.1. Problem solving

Problem solving is a complex process that requires a variety of (meta-)cognitive, affective, and conative processes (Jonassen, 2000). It is often defined as "cognitive processing directed at transforming a given situation into a goal situation when no obvious method of solution is available to the problem solver" (Mayer, 1994, p. 248). According to Mayer, problem solving takes place in the solver's mind and can only be inferred from the solver's behaviour. Other authors argue that the solver's behaviour is not only a proxy of their internal problem-solving processes, but the very essence of problem solving itself (Skinner, 1953). Wheatley and Wheatley (1984) conceptualise problem solving as "What (...) you do when you don't know what to do" (p. 22). In many cases, both internal processes and physical actions (non-verbal and verbal) are needed to solve real-world problems. From the perspectives of embedded and embodied cognition, cognitive, physical, and environmental resources should be in concert with each other to efficiently solve problems (Pouw et al., 2014). In addition, the willingness to engage in problem solving and to persist when faced with difficulties are critical components of the problem-solving process (Jonassen, 2010). In this study, we therefore conceptualise problem solving as the dynamic interplay between cognitive, metacognitive, physical, volitional, and social processes involved in resolving an unfamiliar task or situation.

Following Dewey (1910) and Polya (1945), many scholars have decomposed the problem-solving process into separate steps or phases. Most problem-solving phase models address mathematical problems (for an over view see: Rott et al., 2021; Salminen-Saari et al., 2021). In contrast to these domain-specific models, the PISA model is designed to measure cross-disciplinary problem-solving competencies (OECD, 2004, 2013). The PISA model addresses real-world problems that do not rely on domain-specific knowledge and skills. The PISA model is a more fine-grained elaboration of Polya's four phase model. It is particularly suitable for assessing problem solving in young children because it emphasises exploratory and creative processes that are essential for young children's successful problem solving (Evans et al., 2021). Moreover, the PISA framework's focus on authentic, cross-disciplinary problems aligns well with

**Table 1**

Eight steps of the PISA problem-solving framework (adapted from OECD, 2013).

| Step          | Description  |
|---------------|--|
| Exploring     | Observing and interacting with the problem, searching information, and finding limitation and obstacles                                  |
| Understanding | Understanding given and obtained information and demonstrating understanding of relevant concepts  |
| Representing  | Building a mental representation and constructing tabular, graphical, symbolic or verbal representations                                 |
| Formulating   | Generating hypothesis by identifying relevant elements of the problem and their interrelationships and critically evaluating information |
| Planning      | Setting and clarifying (sub-)goals and devising a plan or strategy to reach that goal  |
| Executing     | Carrying out a plan  |
| Monitoring    | Monitoring progress towards the goal, checking (intermediate) results, identifying unexpected events, and making adjustments as required |
| Reflecting    | Critically evaluating solutions from different perspectives, identifying additional information needs, and communicating progress        |

the goals of early STEM education (Moore et al., 2020). Table 1 presents the eight steps of the PISA problem-solving framework (OECD, 2013)

### 1.2. Observing problem-solving behaviours in preschoolers

Problem-solving behaviours can be defined as any observable, physical, verbal, or non-verbal activity that is triggered by the solver's intention to solve a problem (Gunawan et al., 2019; Uher, 2016). Systematically observing preschoolers' behaviours provides a window through which internal cognitive processes can be inferred (Mayer, 1994). It also provides educators with opportunities to assess preschooler's problem-solving competencies to improve and tailor their learning activities to fit the children's needs (Brassard & Boehm, 2007). Ideally, assessment of preschoolers' competencies is embedded in authentic, play-based activities (DeLuca et al., 2020; Geist, 2025).

Playful STEM activities are particularly suitable to observe and develop problem-solving competencies in preschoolers (Geist, 2025; MacDonald et al., 2022). STEM is the interdisciplinary approach to learning and teaching science, technology, engineering, and mathematics (Johnson et al., 2020). Problem solving is inherent to all four STEM disciplines (Priemer et al., 2020) and is considered "the heart of STEM education" (Dubosarsky et al., 2018, p. 259). Engineering design is "the interdisciplinary glue" in STEM (Tank et al., 2018, p. 183). It is a creative process in which the engineer tries to craft or construct a solution that requires the application of mathematical and scientific principles and the use of technical devices. Engineering design problems are typically authentic, hands-on, ill-structured problems with multiple solutions, including specific constraints and success criteria (Arik & Topçu, 2022; Cunningham et al., 2018; Purzer & Douglas, 2018). Engineering design problems tap into children's natural curiosity and urge to explore the world around them (Isabelle et al., 2021; Moore et al., 2018). These problems also share many similarities with real-life situations during children's construction play (Shechter et al., 2021). Therefore, Gold et al. (2021) argue that observing children's engineering play is an integrative way to assess their cognitive abilities. Despite its potential, engineering has long been underrepresented in early STEM education (Bustamante et al., 2018; English, 2018).

The engineering design process parallels the problem-solving process in several ways, as it starts with establishing a problem or a need and aims to find a solution that meets certain criteria under certain constraints (Cunningham et al., 2018; Dubosarsky et al., 2018; Gold et al., 2021). According to Isabelle et al. (2021), the engineering design process provides a framework to help children develop their problem-solving skills. Recent studies revealed that preschoolers spontaneously engage in engineering-based problem solving during playful activities (e.g., Lippard et al., 2019; Yuan et al., 2024). Shechter et al. (2021) observed how 4- to 6-year-olds brainstormed, planned, tested, and improved their solution to a bridge building task.

In previous studies, various observation instruments and coding schemes have been developed to map aspects of preschool-aged children's problem-solving processes in STEM. Escolano-Pérez et al. (2019) created a coding scheme to capture 5-year-olds' planning, monitoring, and evaluating skills during solving a tangram puzzle. Bryce and Whitebread (2012) asked 5-7-year-olds to build a train track in a specified shape and observed planning, monitoring, and evaluating behaviours. Planning, monitoring, and evaluating behaviours are also at the heart of the MetaSCoPE coding scheme used by Marulis and Nelson (2021) to observe 3-5-year-olds building complex structures from multidimensional Wedgits blocks according to a depicted model. These three coding schemes make use of well-structured puzzle-problems and are embedded in research on metacognition, rather than problem solving, which explains the focus on monitoring and control.

Other coding schemes focus on problem solving within an engineering context. The Early EHoM Coding Scheme created by Shechter et al. (2021) addresses engineering habits of mind of 5-6-year-olds during a well-structured, open-ended bridge-building task, including elements related to understanding, formulating, planning, executing, and monitoring. Although Engineering Habits of Mind (EHoM) shares many similarities with problem solving, the EHoM Coding Scheme is specifically designed for this bridge-building task, making it difficult to apply the coding scheme to other tasks. The Engineering Preschool Children Observation Tool (EPCOT), developed by Anggoro et al. (2021), captures the problem-solving behaviours of 3-5-year-olds during STEM-related classroom activities including behaviours related to exploring, understanding, formulating, planning, executing, and monitoring. As one of the aims of EPCOT was to evaluate the use of vocabulary related to the engineering design process, it largely focuses on oral expression (e.g., repeat, describe, explain, articulate, propose, predict, discuss, state). In addition, the coding scheme largely focuses on knowledge about scientific phenomenon.

None of the above-mentioned studies aimed to develop an observation scheme that exclusively and systematically covers all the activities involved in the problem-solving process. For all mentioned instruments, intercoder reliability was investigated. Some studies also investigated correlations between codes (Shechter et al., 2021) or with other measures (Bryce & Whitebread, 2012; Marulis & Nelson, 2021). But none of the studies provide further elaboration on the psychometric properties of the observation scheme. This same lack of psychometric soundness has been identified by Lippard et al. (2017) as a major obstacle in early engineering research. Scrutinising critical aspects of the validity and reliability of an observational instrument is crucial to its use in research and practice. Therefore, there remains a need for a comprehensive observation instrument that researchers can use to investigate problem-solving behaviours in preschoolers across various, authentic contexts, with clearly established psychometric properties for key aspects of validity and reliability.

### 1.3. The present study

This study is part of a larger research project that investigates how assessment for learning can support the development of problem-solving competencies in preschoolers (ISCED 02; UNESCO Institute for Statistics, 2012). To effectively assess and monitor this

development, there is a clear need for comprehensive, psychometrically sound assessment methods that can be used repeatedly in classroom settings.

The present study aims to develop and establish the psychometric properties of the Problem-solving Observation Scheme for Playful Engineering and the Real-world (PROSPER), a comprehensive observation instrument for systematically observing preschoolers' problem-solving behaviours. PROSPER operationalises the PISA problem-solving framework for preschool-aged children by mapping and describing observable behaviours across the different steps of the problem-solving process. It offers researchers and educators valuable insights into how preschool children approach and solve ill-structured, real-world problems. This study elaborates on the development of the instrument and the establishment of structural validity (construct validity), content validity, ecological validity, intercoder reliability, and across-task consistency (generalisability).

## 2. Method

### 2.1. Design

PROSPER was developed based on videotaped observations of preschoolers solving three different engineering design problems. Five researchers and five teachers identified problem-solving behaviours in the recordings. The identified problem-solving behaviours were then coded and categorised according to the steps of the PISA framework. Finally, psychometric properties of the PROSPER were assessed.

### 2.2. Participants

Participants were 12 children (5 girls and 7 boys) between 4 and 6 years old (mean age = 66 months, SD = 5.45) enrolled in a Flemish (Belgium) preschool. Only children whose parents had given written consent were included. Parents were also asked to provide information on the children's gender, age, and linguistic background. Alongside the language spoken at school, four children spoke another language at home (2 English, 1 Arabic, 1 Polish).

The five authors identified problem-solving behaviours from the recordings. Additionally, five preschool teachers (all female) with 6 to 41 years of experience participated in identifying problem-solving behaviours from the recordings.

### 2.3. Data collection

#### 2.3.1. Engineering design problems

Three engineering design problems were developed for this study. The Volcano Bridge Problem required participants to build a bridge over a papier-mâché volcano with only beer mats and toilet rolls (inspired by Shechter et al., 2021). In the Hook-a-Duck Problem, participants were asked to create a tool to get a ringed rubber duckling out of a transparent tube with functional and non-functional craft materials, cutlery, and tools (inspired by Evans et al., 2021). The Storm Problem required participants to build a stable for two sheep that can withstand the wind of a hairdryer using only natural materials (adapted from Thunder et al., 2022). Examples of solutions to the problems can be found in Appendix A.

The three problems were purposefully selected and developed to be solved by building or crafting with readily available materials, to include three constraints or success criteria, and to allow for multiple solutions and solution paths (Cunningham et al., 2018; Dubosarsky et al., 2018). We assured that the three problems were diverse, enjoyable, required no prior domain-specific knowledge, and presented an appropriate challenge for preschoolers, aligned with their "zone of proximal development" (Vygotsky, 1978). The problems were designed based on the design parameters provided by Cunningham et al. (2018), discussed with preschool teachers, pilot tested, and fine-tuned with five children aged 4 to 6 to ensure that all children can get started and get stuck (low threshold, high ceiling. See: NRICH, 2013). To ensure ecological validity, the problems closely resembled typical preschool classroom activities. All details about the problems are available on the Open Science Framework (OSF; [https://osf.io/k6x5r/?view\\_only=da1682d0dbd94153b70c10ef188a7f3bdoi:10.17605/OSF.IO/K6X5R](https://osf.io/k6x5r/?view_only=da1682d0dbd94153b70c10ef188a7f3bdoi:10.17605/OSF.IO/K6X5R)).

#### 2.3.2. Procedure

All participants solved the problems individually in a spare classroom at the participants' school during school hours. Participants were presented with the three problems at one-week intervals. No time limit was set to solve the problem. All data were collected by the first author.

After demonstrating the technical equipment, the participants were invited to explore the available materials and were asked to name all the objects. Next, the problem, the goal state, and success criteria were explained and visually demonstrated with assistance of Mr Rabbit, a soft bunny hand puppet. The hand puppet was used to put the children at ease, to motivate them, and to elicit talk (Dorie et al., 2013). Mr Rabbit acted as a naïve 'alien' puppet who doesn't know how to solve the problem and asks for the child's help.

During the actual problem-solving process, interventions were kept to a minimum. The researcher only intervened when the participants lost engagement with the assignment, did not act in accordance with the assignment (e.g., participants violated constraints multiple times), asked for help, got really stuck (e.g., participants repeated the same action several times without success, or indicated that they really did not know what to do), or showed clear signs of upcoming frustration (e.g., participants became angry or sad). In these cases, the researcher followed a predefined protocol, responding with prompts that became progressively more concrete and directive. Prompts were given as questions (Wood et al., 1976). Only when this was not sufficient to overcome the impasse, more direct

support was offered by demonstrating part of a potential solution.

Once a solution was found, the participants were asked to recall the problem and reflect on the outcome and the process (Anggoro et al., 2021; Epstein, 2003; Stoll et al., 2012), guided by Mr Rabbit. As in the Dorie et al. (2013) study, the puppet was asleep during the process and didn't know what had happened. This encouraged children to explain what they did and what happened to 'someone' who was not present. The reflective questions gradually increased in difficulty from simple closed questions (e.g., "Was it difficult?") to more complex open questions (e.g., "What was difficult?", "Why didn't it work?", "What else did you try?"). From the second problem onwards, the children were also asked to recall the previous problem(s) and solution(s) before the new problem was explained. The observations lasted on average 29 minutes (range = 12–47 min, SD = 9 min) resulting in 17.33 hours of footage. The full protocol can be found on OSF (doi:10.17605/OSF.IO/K6X5R).

### 2.3.3. Recording materials

All sessions were recorded with two cameras (Logitech C922 Pro Stream Webcam) which captured the working space from two angles. The inbuilt omnidirectional microphone of one of the camera's was used to register all speech and sounds. Footage from both cameras was processed by the open-source video recording application OBS Studio (version 30.2.2, www.obsproject.com).

## 2.4. Developing PROSPER

Developing PROSPER involved several stages. First, a preliminary observation scheme was drafted based on the existing scientific literature (e.g., Anggoro et al., 2021; Bryce & Whitebread, 2012; Marulis & Nelson, 2021; Shechter et al., 2021) and open narrative descriptions made during the initial screening of recordings. Second, problem-solving behaviours were identified from the recordings and briefly described. Then, these chunks of identified problem-solving behaviours were coded in several rounds (Saldaña, 2013). Figure 1 illustrates the different stages of the developmental process of PROSPER.

As problem solving is a process, the eight steps of the PISA problem-solving framework were conceived as the main categories and the specific behaviours as the codes (Saldaña, 2013). As the codes refer to observable actions, gerunds (the "-ing" form of verbs) were used (Thornberg & Charmaz, 2014). The observation scheme was developed iteratively. Over the first three rounds of coding, codes were created, removed, moved, merged, and split (Thornberg & Charmaz, 2014). The final round of coding was used to establish intercoder reliability. This section outlines the development of PROSPER, focusing on the steps taken to ensure the instrument's comprehensiveness, validity, and reliability.

### 2.4.1. Creating a preliminary observation scheme

The eight steps of the PISA problem-solving framework were used as the main categories in the observation scheme (OECD, 2013).

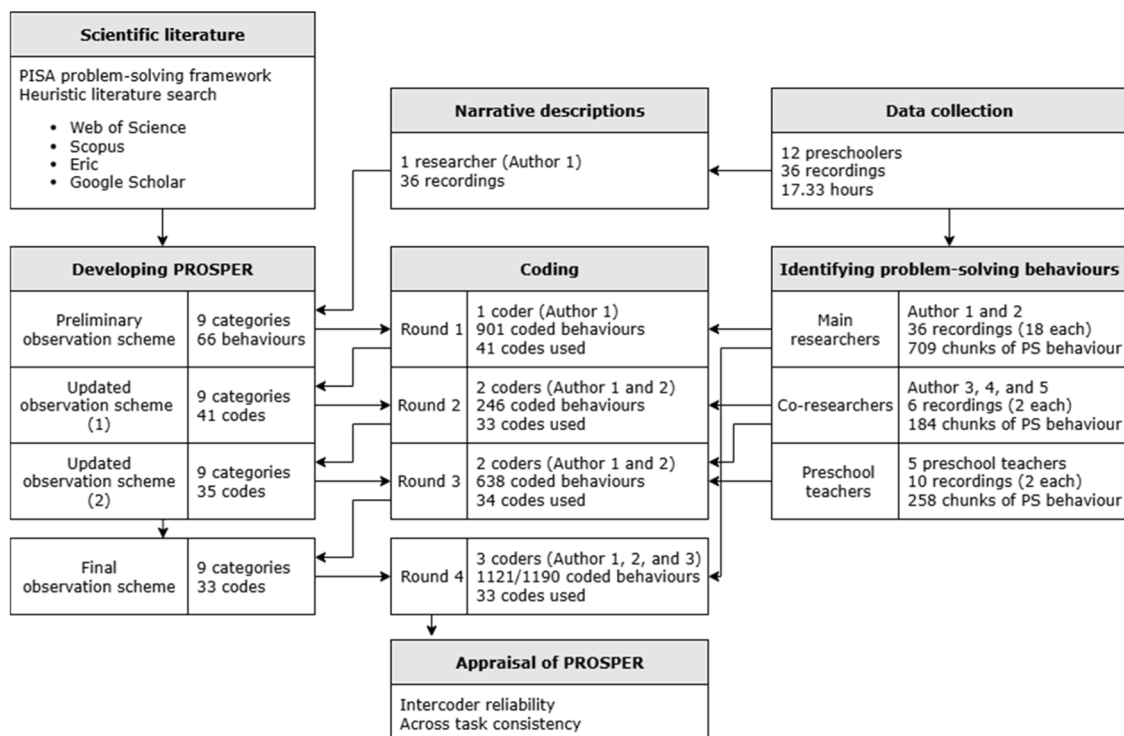


Fig. 1. Developing PROSPER: flowchart.

For every category, related problem-solving behaviours were looked up through heuristic literature search in the databases of Web of Science, Scopus, Eric, and Google Scholar. The list of problem-solving behaviours was further completed based on the field notes taken during data collection and the first author's open, narrative descriptions of all recordings (Saldaña, 2013). All behaviours were categorised across the eight PISA steps. A residual category was created for generic, overarching problem-solving behaviours that could not be assigned to any of the PISA steps.

#### 2.4.2. Identifying problem-solving behaviours

A protocol and a structured observation form for identifying problem-solving behaviours were developed and piloted. The first two authors each watched half of the recordings. Two recordings were processed by both researchers. Doubtful cases were discussed and resolved in consensus to ensure uniform interpretation of the protocol. Problem-solving behaviours were described and categorised according to the PISA steps.

To ensure content validity and to avoid blind spots, the last three authors, and five preschool teachers also identified problem-solving behaviours in the recordings. All observers were trained and were provided with a protocol and a digital and paper version of the observation form. They each processed two different recordings independently. Recordings were purposefully chosen to maximise variation in problems, gender and home language of the preschoolers, and duration of the session. Divergent perspectives were analysed by comparing behaviours identified by the first two authors and the other observers.

#### 2.4.3. Coding problem-solving behaviours

In the first round of coding, the preliminary observation scheme was applied by the first author to code all chunks of problem-solving behaviours identified by the first two authors. The aim of this first round of coding was to identify missing behaviours and to update the preliminary observation scheme. After coding and in consultation with the last two authors, the observation scheme was updated by adding, splitting, merging, moving, and deleting codes.

In Round 2, the updated observation scheme was used to code the behaviours identified by the three last authors. In this round, the first two authors independently coded all chunks of identified problem-solving behaviours. If the chunks described more than one action or behaviour, multiple codes were assigned, resulting in as many coded behaviours as actions described in the chunks of identified behaviours.

After coding, intercoder reliability was assessed by calculating percentage of agreement, Cohen's kappa based on pairwise deletion, and Cohen's kappa with missing values as a regular category (De Raadt et al., 2019). When one of the coders assigned more codes to a chunk, missing values were produced for the other. Cohen's kappa with missing values as a regular category is considered the lower bound as it is heavily biased downward (De Raadt et al., 2019). We aimed to achieve a percentage of agreement > 80% and a strong intercoder agreement with  $\kappa > .80$  as proposed by McHugh (2012) and a percentage of agreement > 80%. Intercoder reliability was assessed on the level of the categories as well as on the level of the codes. Intercoder reliability statistics are presented in Table 2 (Section 3.2) and discussed in Section 3.3. Differences between the coders were discussed between the two coders. Again, the observation scheme was adjusted by adding, splitting, merging, moving, or deleting codes.

In Round 3, the updated observation scheme was used by the first two authors to code all chunks of identified behaviours by the three last authors and the preschool teachers. Once again, intercoder reliability was assessed, discrepancies were discussed between the first two authors, and the observation scheme was adjusted. This resulted in the final observation scheme (PROSPER).

#### 2.4.4. Psychometric properties of PROSPER

The psychometric properties being addressed for PROSPER include structural validity (construct validity), content validity, ecological validity, intercoder reliability, and across task consistency (generalisability). Structural validity, as part of construct validity, refers to the degree to which the observation scheme reflects the theoretical structure of the construct under investigation (Mokkink et al., 2010). Structural validity was established by grounding the observation scheme in the PISA problem-solving framework (OECD, 2013). As many other problem-solving models found in the scientific literature, the PISA model builds on the

**Table 2**

Summary of the four rounds of coding with intercoder reliability statistics on the level of the codes and on the level of the categories (between brackets).

| Dataset (Behaviours identified by) | Round 1<br>Author 1 & 2 | Round 2<br>Author 3, 4, & 5 | Round 3<br>Author 3, 4, & 5 + teachers | Round 4a<br>Author 1 & 2 | Round 4b<br>Author 1 & 2 |
|------------------------------------|-------------------------|-----------------------------|--|--------------------------|--------------------------|
| Coders (N)                         | Author 1                | Author 1 & 2                | Author 1 & 2                           | Author 1 & 2             | Author 1 & 3             |
| Codes used (N)                     | 41                      | 33                          | 34                                     | 33                       | 33                       |
| Coded chunks (N)                   | 709                     | 184                         | 441                                    | 709                      | 709                      |
| Coded behaviours (N)               | 901                     | 246                         | 638                                    | 1121                     | 1190                     |
| Missing values                     | 25                      | 88                          | 145                                    | 245                      | 290                      |
| Percentage of agreement            | NR                      | 86.71% (92.05%)             | 84.18% (88.57%)                        | 89.84% (92.12%)          | 82.00% (86.50%)          |
| $\kappa$ pairwise deletion         | NR                      | .856<br>(.893)              | .827<br>(.858)                         | .893<br>(.907)           | .810<br>(.841)           |
| $\kappa$ regular category          | NR                      | .553<br>(.714)              | .664<br>(.723)                         | .711<br>(.706)           | .624<br>(.632)           |

Note. NR: Not Relevant;  $\kappa$ : Cohen's kappa.

well-established phase model of Polya (1945).

Content validity refers to the degree to which the content of an instrument reflects the construct (Terwee et al., 2018). Content validity was ensured by aligning PROSPER with the eight steps of the PISA problem-solving framework (OECD, 2013) and by having five experienced preschool teachers observe two video recordings each and identify problem-solving behaviours. The involvement of practice experts ensured that the behaviours included in the observation scheme were relevant, representative, and meaningful within the preschool context (Boateng et al., 2018).

Ecological validity is part of external validity and refers to the degree to which results can be generalised to real-world contexts. Ecological validity was ensured by creating problems closely related to preschoolers' play activities, by using materials familiar to the participants, by mimicking authentic teacher-child interactions, by conducting the sessions in a familiar classroom environment at the participants' own school, and by involving teachers in the identification of problem-solving behaviours (Rhodes et al., 2024).

Intercoder reliability refers to the consistency between independent coders. This was established by having the first two authors independently code the problem-solving behaviours they had previously identified using the final observation scheme (Round 4). As the two coders had developed the observation scheme and were familiar with the data, a third coder (Author 3), who was not involved in developing the observation scheme and was unfamiliar with the data, coded the same dataset. Intercoder reliability was assessed by calculating percentage of agreement and Cohen's kappa for coding conducted between Author 1 and Author 2, and between Author 1 and Author 3. As before, we aimed to achieve a strong intercoder agreement with  $\kappa > .80$  (McHugh, 2012) and a percentage of agreement  $> 80\%$ . Intercoder reliability statistics of the final observation scheme are presented in Table 2 (Section 3.2) and discussed in Section 3.4.

Finally, across-task consistency or task dependency was assessed to determine whether the three problems elicited comparable levels of problem-solving behaviours. This was achieved by comparing the frequencies of coded problem-solving behaviours and their corresponding categories assigned by the first author in Round 4.

Quantitative analyses were conducted in R (version 4.3.2, R Core Team, 2023). The package *irr* (version 0.84.1, Gamer et al., 2019) was used to calculate intercoder reliability measures. All protocols are available on Open Science Framework ([https://osf.io/k6x5r/?view\\_only=da1682d0dbd94153b70c10ef188a7f3b](https://osf.io/k6x5r/?view_only=da1682d0dbd94153b70c10ef188a7f3b) doi:10.17605/OSF.IO/K6X5R).

### 3. Results

#### 3.1. A preliminary observation scheme

Based on the scientific literature and the narrative descriptions of the recordings, a list of 66 problem-solving behaviours was composed by the first author. The list consisted of 46 behaviours categorised in the eight PISA steps. Twenty behaviours were initially not categorised in one of the eight PISA categories and related to identifying problems, goal setting, decision making, engagement, persistence, autonomy, self-efficacy, motor skills.

#### 3.2. Identifying problem-solving behaviours

Together, the first two authors identified and described 709 chunks of problem-solving behaviours in the 36 recordings (see Table 2). On average, 19.69 behaviours were identified per recording with a range of 15 to 27 ( $SD = 3.28$ ). The last three authors identified 184 problem-solving behaviours in 6 recordings with an average of 30.67 described behaviours ( $SD = 15$ ,  $Min = 18$ ,  $Max = 60$ ). The preschool teachers identified 258 behaviours in 10 recordings with an average of 25.8 described behaviours per recording ( $SD = 7.8$ ,  $Min = 14$ ,  $Max = 41$ ). Sometimes all behaviours were neatly recorded separately as intended, sometimes detailed descriptions of sequences of actions combining several behaviours were provided. Two examples illustrate the differences in the registration of identified problem-solving behaviours between observers. The numbers between square brackets indicate the time stamp of the registered behaviour. Between brackets we indicate whether the behaviours were identified by a researcher or by a teacher, the observation number, and the problem.

[33:25] "Tests kitten on bridge."

[34:00] "Puts card on toilet roll (Trial and error)."

[34:34] "Tests kitten on bridge."

[35:55] "Another toilet roll."

[36:00] "Tests kitten on bridge."

[36:10] "Places card on bridge."

[36:55] "Tests kitten on bridge."

[37:27] "Evaluates that this is a good solution." (Researcher, Observation 16, Volcano Bridge Problem)

[37:30] "Starts for the third time. Makes another attempt. Rebuilds the same tower again. Now places a beer coaster on top so the cat can sit on it. It falls over... Persists but does not come to new experiments or plans. Keeps making the same mistakes". (Teacher, Observation 18, Volcano Bridge Problem)

A rich variety of problem-solving behaviours was observed, covering seven of the eight PISA steps. In this section, problem-solving behaviours are described and illustrated with examples. Exploring is the first step in the PISA framework and refers to deliberately gathering information about the problem, materials, and possibilities by observation and interaction with the problem and environment. Participants explored by observing, manipulating, describing, and comparing materials and associating them with other materials, potential application, or past experiences. We also observed children rummaging through provided materials to see what they can use and asking questions for more information. However, some children needed encouragement to explore.

*"Names objects; describes if she does not know the name (bark); Associates (straw with cowboy); manipulates only some objects."* (Researcher, Observation 31, Storm)

*"Stares at the box of materials without touching or saying anything. Needs encouragement to explore."* (Researcher, Observation 3, Volcano Bridge)

Demonstrating understanding of the problem, its key elements, constraints and success criteria is the second step. Although understanding is a mental process, we identified behaviours from which understanding can be inferred. We found that some participants listened and watched attentively during the problem description while others were already up and running before the problem was even properly explained. Children also confirmed understanding by nodding or verbal expression and by repeating or demonstrating the problem and criteria. Finally, understanding could be derived from the extent to which the children acted in accordance with the criteria.

*"Listens and watches Mr Rabbit and Kitten/Vulcano; Gives verbal affirmation or nods to questions; Touches the volcano and pretends it burns him (demonstrates criteria); Can repeat and demonstrate criteria."* (Researcher, Observation 17, Volcano Bridge)

*"Indicates he is ready and demonstrates how the kitten is shot over the volcano using his canon. This is, of course, not in accordance with the assignment."* (Researcher, Observation 34, Volcano Bridge)

Representing, creating a mental, verbal, visual, or physical representation of the problem or goal, was difficult to observe. The few behaviours that could possibly be attributed to representing, were more associated with demonstrating understanding, as in the excerpt below.

*"Demonstrates what the kitten wants to do by going over the volcano with the kitten in the air."* (Researcher, Observation 15, Volcano Bridge)

Formulating refers to coming up with possible ways to solve the problem and evaluating alternatives. Generating new ideas can be observed when preschoolers verbalise their ideas or when preschoolers try out a new solution. While children frequently generated new ideas throughout the entire problem-solving process, considering was only observed to a limited extent.

*"Suggests what he can try (sticking tape to the knife, piercing [a skewer] at the bottom of the duckling, using tape to attach a pair of scissors 'to that', tying pipe cleaner around the stick) and demonstrates this with the materials or gestures".* (Researcher, Observation 7, Hook-a-Duck)

*"Says 'I have an idea' (sucking up the kitten), evaluates it and concludes it will not work."* (Researcher, Observation 15, Volcano Bridge)

Planning behaviours involve preparing for and anticipating actions. As with representing, planning is often a mental process. However, we observed some behaviours that reflect planning such as collecting materials, clearing the workspace, and demonstrating or verbalising next steps.

*"Prepares, lays out the collected materials and sets his goal 'Attaching to each other'."* (Teacher, Observation 28, Hook-a-Duck)

*"'First I have to make a wall ... And then with a stick ....'"* (Researcher, Observation 12, Storm)

Executing refers to working towards a solution or trying to do so. Executing is about doing things and doing things well. Besides building and crafting solutions, we also saw preschoolers refining their solutions. Sometimes preschoolers lacked a hand or motor skills to perform a desired action and asked for help.

*"Tries; trial & error with different materials."* (Researcher, Observation 27, Hook-a-Duck)

*"Tinkers. Tries to make various things, but with no real clear purpose."* (Researcher, Observation 6, Hook-a-Duck)

*"Tries first by herself and then asks the teacher to help her tear the tape; tries again a 2nd time first by herself; 3rd time she immediately asks for help; 4th time she succeeds by herself!"* (Researcher, Observation 23, Hook-a-Duck)

Monitoring means keeping track of progress, checking in between to see if the criteria are being met, and changing plans if necessary. We observed children testing, checking, measuring, and counting. Children established additional problems, adjusted elements of the solution, and indicated that a solution had been reached.

*"Spontaneously tests with the hairdryer and says, 'Here comes the storm again.' One stone falls over. He says, 'Not such a sturdy little house.' But then adds, 'These ones work well,' referring to the two stones that stayed up. 'But this stone doesn't work,' and puts the stone aside."* (Researcher, Observation 32, Storm)

*“Shouts ‘Finished!’, but she has NOT tested the criteria: her house has only two walls, the sheep are elsewhere, and she hasn’t tested with the hairdryer.” (Researcher, Observation 10, Storm)*

Reflecting refers to evaluating the solution and the process. To elicit reflective thoughts, the reflection phase was guided by the researcher by means of Mr Rabbit. Children were asked to recall elements of the process and evaluate elements of the process and the solution. Some children were able to recall and evaluate several attempts, others struggled with it or said things that did not reflect the reality. When asked if they enjoyed doing it, almost all the children answered yes, even if this does not always seem consistent with their expressions during the process. When the researcher inquired further, some children were able to identify situations they found unpleasant.

*“Tells Mr Rabbit it’s a good house because it doesn’t blow away. He checks with the hairdryer. He also indicates that it has a roof, three walls and the sheep can get out easily.” (Researcher, Observation 14, Storm)*

*“Recalls several failed trials and explains why it failed (e.g. the pencil was too short).” (Researcher, Observation 4, Hook-a-Duck)*

*“Says it was not difficult, but he needed help with everything.” (Researcher, Observation, Storm)*

*“Indicates that he liked it, but on the video it seems to be different.” (Researcher, Observation 11, Storm)*

*“Says she didn’t like the bridge collapsing that much.” (Researcher, Observation 2, Volcano Bridge)*

The observers also identified some generic behaviours that span the whole problem-solving process. Although participants often were focused on solving the problem, some children quickly became tired of it, started playing, or exhibited disruptive behaviour (commitment). Similarly, some children persisted after several failed attempts, while others gave up quickly or got frustrated (perseverance). The observers noted that some participants took ownership of the problem and worked autonomously without needing or asking for support, while others relied heavily on the researcher for support (autonomy). Finally, the observers noted how the children responded to the researchers’ prompts, which could either be to follow the prompt, ignore the prompt, or respond to the prompt but do something else than what was suggested (responsiveness).

*Commitment: “Starts playing/throwing the kitten. Loses engagement.” (Researcher, Observation, Volcano Bridge)*

*Perseverance: “Gets visibly frustrated (hits around, turns away) when rolls keep falling: ‘Aaarg! I can’t do this.’” (Researcher, Observation 36, Volcano Bridge)*

*Autonomy: “Repeats that he does not know how to do it and keeps repeating this when investigator invites (several times) to think of other things; does nothing, waits.” (Researcher, Observation 8, Storm)*

*Responsiveness: “Ignores the researcher’s suggestion to make the house smaller.” (Teacher, Observation 30, Storm)*

### 3.3. Coding problem-solving behaviours

After identifying problem-solving behaviours in the recordings, the chunks of identified behaviours (as the excerpts above) were coded. After the first round of coding, in which the first author coded the behaviours identified by the first two authors, the preliminary observation scheme was discussed within the research team and adapted, resulting in 41 codes. The observation scheme was adjusted by rewording code names (N = 11), merging codes (N = 30), relocating codes to other categories (N = 8), or removing codes (N = 14). Codes were reworded to active verbs (e.g., “Associates materials with other things or situations” became “Associating”). Codes were merged when they were difficult to distinguish (e.g., “Touches materials” and “Makes movements with the materials” were merged into “Manipulating”) or when they related to similar processes (e.g., “Thinking ahead” and “Sequencing”). Codes were relocated to another category when they turned out to be more related to another category. Codes were removed when they turned out to be difficult to observe or interpret (e.g., “Thinks”).

For the second round of coding, the identified behaviours by the last three authors were used. In this round, 184 chunks of identified behaviours were coded, resulting in 246 coded behaviours. All metrics are presented in [Table 2 \(Section 3.2\)](#). Strong intercoder agreement was established with 86.71% agreement between both coders and Cohen’s kappa with pairwise deletion .856 ( $p < .001$ ). Cohen’s kappa with missing values as a regular category was below our threshold ( $\kappa = .553$ ,  $p < .001$ ). In many chunks, multiple behaviours were described, resulting in multiple codes per chunk. When both coders assigned a different number of codes to a chunk, missing values were produced for the other coder. Therefore, we clarified the guidelines to deal with multiple behaviours in one chunk. In addition, the observation scheme was discussed within the research team and adapted. Three codes were adjusted, six were merged, one was split, and three were removed. We also added a code for responsiveness to the category ‘Generic’.

For the third round of coding the first two authors coded 441 chunks of behaviour identified by the last three authors and the teachers, resulting in 638 coded behaviours. Agreement between both raters was comparable with Round 2 (percentage of agreement = 84.18%;  $\kappa$  pairwise deletion = .827,  $p < .001$ ). Cohen’s kappa with missing values as a regular category improved up to .664 ( $p < .001$ ). Once again, the observation scheme was discussed within the research team and adapted. We adjusted one code, merged seven codes, and added three new codes: complying, asking ideas (which was a spin-off from the more general asking help), and declaring that a solution has been found.

### 3.4. Psychometric properties of PROSPER

The final observation scheme (PROSPER), including definitions and examples, is presented in [Appendix B](#). PROSPER consists of nine categories representing the eight steps of the PISA framework and a generic category for overarching behaviours. Thirty-three behaviours were assigned to seven of the eight PISA categories: exploring (8 behaviours), understanding (4 behaviours), formulating (3 behaviours), planning (2 behaviours), executing (3 behaviours), monitoring (6 behaviours), and reflecting (3 behaviours) and four overarching behaviours were assigned to the generic category: commitment, perseverance, autonomy, and responsiveness. No behaviours were assigned to the step of representing (see PISA framework) because we were not able to identify representing behaviours that could not equally well be attributed to one of the other categories.

#### 3.4.1. Intercoder reliability

The results of the intercoder reliability analysis of the final observation scheme are presented in [Table 2](#) ([Section 3.2](#)). Percentage of agreement (89.84%) and intercoder reliability ( $\kappa_{\text{pairwise deletion}} = .893$ ,  $p < .001$ ) between the first two authors who developed the observation scheme was strong. Cohen’s kappa with missing values as regular category was still good ( $\kappa_{\text{regular category}} = .711$ ,  $p < .001$ ). Here we recall that missing values are produced when two coders assign a different number of codes to a chunk of identified behaviours. Sometimes it was difficult to assess whether one or more behaviours were described, especially when generic behaviours were involved. For example, “Tests by taking the sheep out of the house once (at teacher’s suggestion).” (Researcher, Observation 8, Storm) was coded ‘testing’ by Author 1 and ‘testing’ and ‘responsiveness’ by Author 2.

Although slightly less, the percentage of agreement (82.00%) and intercoder reliability ( $\kappa_{\text{pairwise deletion}} = .810$ ,  $p < .001$ ) between Author 1 and Author 3, who was not involved in developing the observation scheme, was still strong. This indicates that other coders can reliably use the observation scheme to code problem-solving behaviours.

Most discrepancies between coders relate to the code doing ( $N_{\text{Discrepancies}} = 34$ ). Doing is defined as working towards a solution through physical actions. Therefore, it is sometimes difficult to differentiate doing from testing ( $N_{\text{Discrepancies}} = 9$ ) or modifying ( $N_{\text{Discrepancies}} = 7$ ). For example, “Tries to fish out the duckling with a long dessert spoon.” (Researcher, Observation 4, Hook-a-Duck) was coded ‘doing’ by Author 1 and ‘testing’ by Author 2.

#### 3.4.2. Across-task consistency

[Table 3](#) shows the number of coded behaviours for each problem and category based on the codes provided by the first author in coding Round 4. Although the numbers in the table should be interpreted with some caution, we note that an equivalent number of behaviours was identified for all three problems. However, we observed some differences between the three problems. The Volcano Bridge Problem elicited least exploring behaviours and most formulating behaviours. To solve the Volcano Bridge Problem, the participants could only use uniform toilet rolls and beer mats, leaving little to explore. On the other hand, participants tried different constructions, which revealed a lot of formulating behaviours. During the Hook-a-Duck problem, preschoolers revealed most exploring behaviours and least planning behaviours. Here, participants were challenged to explore materials by providing a small toolbox filled with many materials in different sizes and shapes. Planning was difficult to observe since most children tried to solve the problem by trial and error. In the storm problem, preschoolers demonstrated most monitoring behaviours and least formulating behaviours. We observed participants counting (e.g., the number of walls), measuring (e.g., whether the walls were sufficiently high), checking stability (e.g., blowing with the hairdryer), and adjusting (e.g., increasing the height of the walls). As most participants continued to build on their initial structure, there was less incentive to come up with or consider new ideas.

## 4. Discussion

Studying the development of young children’s problem-solving competencies requires comprehensive, valid, and reliable instruments to observe problem solving in an authentic context. To date, such an instrument is lacking. The aim of the present study was to address this gap by developing a comprehensive, valid, and reliable instrument to make preschoolers’ problem solving visible in an

**Table 3**  
Number (and column percentages) of coded behaviours per problem and per category.

| Category      | Volcano Bridge | Hook-a-Duck | Storm       | Total        |
|---------------|----------------|-------------|-------------|--------------|
| Exploring     | 46 (13.98%)    | 89 (26.25%) | 53 (16.51%) | 188 (19.01%) |
| Understanding | 54 (16.41%)    | 36 (10.62%) | 40 (12.46%) | 130 (13.14%) |
| Representing  | 0 (0%)         | 0 (0%)      | 0 (0%)      | 0 (0%)       |
| Formulating   | 37 (11.25%)    | 26 (7.67%)  | 20 (6.23%)  | 83 (8.39%)   |
| Planning      | 12 (3.65%)     | 3 (0.88%)   | 15 (4.67%)  | 30 (3.03%)   |
| Executing     | 39 (11.85%)    | 45 (13.27%) | 35 (10.90%) | 119 (12.03%) |
| Monitoring    | 58 (17.63%)    | 72 (21.24%) | 86 (26.79%) | 216 (21.84%) |
| Reflecting    | 46 (13.98%)    | 52 (15.34%) | 39 (12.15%) | 137 (13.85%) |
| Generic       | 37 (11.25%)    | 16 (4.72%)  | 33 (10.28%) | 86 (8.70%)   |
| Total         | 329            | 339         | 321         | 989          |

*Note.* The total number of assigned codes differs from the total number of coded behaviours in [Table 2](#) because we only used the codes from author 1 and we excluded missing values to make the percentages easier to compare.

authentic context. The Problem-solving Observation Scheme for Playful Engineering and the Real-world (PROSPER) is grounded in the PISA framework and was developed through an in-depth analysis of 36 recordings of 4- to 6-year-olds involving five researchers and five preschool teachers. PROSPER is the first validated observation instrument that covers the whole problem-solving process. It consists of 33 behaviours, categorised across seven of the eight PISA steps. In addition, four generic, overarching problem-solving behaviours were identified. Construct and content validity were established by drawing on a longstanding research tradition on problem-solving phase models and involving preschool teachers in the analysis of recordings. Ecological validity was supported by designing problems closely related to children's play and creating a setting that mimicked a classroom corner and authentic teacher-child interactions. Intercoder reliability was strong and across task consistency was good.

In this section, we first discuss the observed behaviours per step of the problem-solving process in relation to existing scientific literature. Then, we delve into the psychometric properties of the instrument, highlighting the strengths and limitations of the study. Finally, we discuss the implications of our findings for research and education.

#### 4.1. Preschoolers' problem-solving behaviours

We observed a rich variety of problem-solving behaviours involving seven of the eight PISA steps. Exploring is the first step in any problem-solving process and is a crucial element of successful problem solving for preschool children (Evans et al., 2021; Fusaro & Smith, 2018). Preschoolers observed, manipulated, described, and compared materials and associated them with other materials or past experiences. This finding is consistent with the observation of Cunningham et al. (2018) that preschool children aged 5 to 6 engage in and benefit from observing, manipulating, and discussing properties of materials.

The second step is understanding the problem. A problem not well understood is impossible to solve (Polya, 1945). However, it can be difficult to determine whether a preschooler understands the problem, criteria, constraints, and prompts. In developing PROSPER, we have worked around this problem by including behaviours that indicate attempts to understand the problem and the instructions. These behaviours include focusing on instructions (listening and confirming), repeating them and acting on them (complying). Although PROSPER can be helpful to identify relevant behaviours related to (attempts to) understanding, making understanding visible remains challenging. Demonstrating understanding requires strong receptive and expressive language skills (Vessonen et al., 2023). Many children experienced difficulties with repeating the criteria and constraints immediately after the problem was stated by the researcher. This raises the question of whether they did not understand the criteria and constraints or just had trouble articulating them. Further research is needed to find ways to make understanding visible. The use of visual aids can be helpful, such as showing children pictures of solutions that either meet or violate the constraints and criteria, and asking them to identify the acceptable and unacceptable solutions.

Step three is representing. In our study, we were unable to make representing visible through observation. First, representing is typically an internal mental process (Mayer, 1994), impeding its operationalisation. In older students, mental representations can be externalised by tabular, graphical, symbolic or verbal representations (OECD, 2013). In preschool, children may be asked to externalise their representation of the problem or solution through drawing and verbal explanation (Bartholomew et al., 2019; Cartwright, 2023). Although 4-6-year-olds can be taught to graphically express their representations (Shiakalli & Zacharos, 2012), both drawing and interpreting their drawings remains challenging for them (Bat Or et al., 2022). Second, distinguishing between behaviours related to representing and understanding proved difficult. Representing and understanding are conceptually distinct yet closely related aspects of the problem-solving process. This is illustrated by Hayes (1981) who defines representing as "understanding the nature of the gap to be crossed." (p.11). Pretz et al. (2003) define representing as "the manner in which the information known about a problem is mentally organized." (p. 6). Whereas understanding is purely about grasping the problem situation, goal situation, criteria, and constraints, representing refers to connecting all the elements into a coherent mental model or problem space (Jonassen, 2010). For ill-structured or real-world problems, integrating all pieces into a coherent representation proves particularly challenging (Pretz et al., 2003). Articulating the relationships between all elements of the problem and problem space is even more demanding. Given preschoolers' difficulties in externalising their problem representations, the strong conceptual relationship between representing and understanding, and the inherent complexity of representing ill-structured, real-world problems, the question arises whether and how representing can and should be distinguished from understanding in preschool-aged children.

Formulating possible solutions is the fourth step of the problem-solving process. Formulating consists of divergent thinking (thinking up) and convergent thinking (considering). Although strongly related, these are two distinct cognitive processes (Guilford, 1967). Generating new ideas can be observed when preschoolers verbalise their ideas or inferred when preschoolers try out a new solution. While children frequently generated new ideas throughout the entire problem-solving process, considering was only observed to a limited extent. Follow-up research can make preschoolers work together to solve to create a context in which preschoolers are encouraged to spontaneously evaluate and weigh up solutions to make decisions together.

Planning is the fifth step of the problem-solving process. We did not observe many planning behaviours. We observed planning behaviours when participants prepared their work (e.g., by collecting materials) or expressed their thoughts by explained next steps (thinking ahead). Since planning is mainly a mental process, we cannot rule out the possibility that preschoolers made plans internally. However, planning takes time and often participants started doing something immediately, even before the problem was fully explained to them. Therefore, we question whether preschoolers do plan in advance. Although previous research points out that children can plan their actions at the age of five (e.g., Warneken et al., 2014), planning behaviours do not always occur spontaneously. Possible explanations can be a lack of prior knowledge, failure to appreciate the importance of planning, or difficulties with remembering and implementing plans (e.g., Zelazo et al., 1997). Hudson et al. (1997) advocate that external support such as reminders can be beneficial to elicit preschoolers' planning activities.

Executing is the sixth step and refers to acting towards a solution. The PISA framework defines executing as “carrying out a plan” (OECD, 2013, p. 126). According to this interpretation, one needs to have a plan prior to executing. Since preschoolers often start without a clear plan, as discussed above, and only start planning on-the-go (English, 1992), less systematic behaviours with the intention of solving the problem were also categorised as executing (e.g., trial and error and tinkering). Executing includes all physical actions to transform the problem state into the goal state. To distinguish between ‘just doing’ and ‘trying to do it well’, we added a code for refining. This covers all behaviours that are intended to refine the solution beyond the formulated criteria (e.g., make the solution more attractive).

Step seven of the PISA framework is monitoring. We identified a variety of monitoring behaviours, including testing, measuring, determining whether something works or not, modifying, and declaring that a solution has been reached. This observation is in line with the study of Bryce and Whitebread (2012) who observed a dozen different monitoring behaviours during a train track building task. English (1992) argues that children’s monitoring behaviour is a compensation for their lack of domain-specific knowledge and initially non-planned, trial-and-error behaviour. As such, monitoring appears to be the central hub through which the problem-solving process unfolds. This contrasts with problem solving in older learners, where planning appears to play a primary role in guiding the process (e.g., Polya, 1945).

The last step is reflecting. PROSPER distinguishes between three reflective actions: recalling (Epstein, 2003), evaluating the result (e.g., Bonner et al., 2021; Stoll et al., 2012; Zimmerman & Moylan, 2009), and evaluating the process (e.g., Anggoro et al., 2021; Epstein, 2003). We found that preschoolers’ reflections were not always consistent with what we observed. At the same time, some preschoolers could recall failed attempts and even elaborated on possible causes of them. Further research can provide deeper insights into preschoolers’ reflective abilities by comparing their reflections with observed behaviours.

Our observations also revealed four generic, overarching behaviours that cut across the whole process and cannot be attributed to any of the PISA steps: commitment, perseverance, autonomy, and responsiveness. Commitment is the willingness to solve the problem and can be inferred from the eagerness to start and focussed on-task behaviour as opposite to swirling around and engaging in off-task behaviour (e.g., Liljedahl, 2016). Without commitment, children demonstrate little problem-solving behaviours. Therefore, commitment is a necessary condition for reliable observation of problem-solving competencies.

In contrast to Liljedahl (2016), we differentiate between commitment (engagement) and perseverance (persistence). While commitment is about the dedication to solve a problem, perseverance is about the ongoing efforts and resilience to overcome obstacles. Perseverance can be derived from children’s reactions to unsuccessful attempts which can be positive (e.g., rebuilding or reconstructing pillars of the bridge after collapse) or negative (e.g., giving up after the first collapse).

Autonomy is the extent to which a child takes ownership on solving the problem. According to Deci and Ryan (1987), autonomy is about choice and feeling responsible (Geist, 2025). It can be derived from the degree to which a child relies on external support. It is important to emphasise that it is not because a child asks for or needs help that they cannot be autonomous. For instance, asking for help in tearing the tape illustrates agency. The same applies to children who need help to overcome an impasse after several failed attempts to solve the problem.

Finally, responsiveness refers to the degree to which a child responds appropriately or in accordance with the support provided (see also McLeod et al., 2024). Responsiveness can be inferred from a change in children’s actions after a prompt is given, such as changing the structure of a building or, conversely, ignoring the prompt. Although there is a paucity of research on children’s responsiveness to teachers’ prompts, it has emerged as an important theme during identification of problem-solving behaviours by both researchers and preschool teachers. We observed some children immediately following up on the suggestions, while others simply ignored the suggestions. Further research is needed to investigate preschoolers’ responsivity to teachers’ prompts in relation to effective and successful problem solving.

Commonly used problem-solving phase models such as those of Polya (1945) and Schoenfeld (1985) don’t cover overarching behaviours that span the entire process. Based on our observations and the identified problem-solving behaviours by the teachers, we argue that generic behaviours such as commitment, perseverance, autonomy, and responsiveness should be part of any instrument that maps preschoolers’ problem-solving behaviours.

In PROSPER, special attention is given to asking questions which is often associated with curiosity (e.g., Fusaro & Smith, 2018; Luce & Hsi, 2015). Young children tend to ask a lot of questions (e.g., Jirout & Klahr, 2020; Wong et al., 2024). For them it is a way to explore, solve problems, and learn. Initially, asking questions was conceived as a separate, generic behaviour. Through iterative coding and discussion of discrepancies within the research team, it emerged that questions had different aims and that they were related to different steps in the problem-solving process. Therefore, asking questions was broken down in asking information (exploring), asking ideas (formulating), asking motor help (executing), and asking confirmation (monitoring). In addition, we observed children who kept on asking for help in only general terms. This indicated a lack of taking responsibility and was integrated in the generic code autonomy.

#### 4.2. Psychometric quality of PROSPER

In developing PROSPER, we ensured strong structural validity, content validity, and ecological validity and assessed intercoder reliability and across task consistency. Construct validity was ensured by firmly rooting the instrument in the PISA problem-solving framework which is in turn embedded in a rich research tradition on problem-solving phase models (OECD, 2013). Content validity was established by involving five researchers and five preschool teachers in identifying problem-solving behaviours. To strengthen ecological validity, three engineering design problems were developed that are closely related to preschoolers’ construction play and crafting activities (e.g., Shechter et al., 2021). Additionally, data were collected in a familiar classroom environment following a protocol that simulated authentic teacher-child interactions. It would be interesting to further validate the instrument with

a larger sample, in other education systems and settings (cross-cultural validity, ecological validity), with other STEM-problems (external validity), against other problem-solving measures (concurrent validity), and to examine how well the instrument predicts later learning outcomes (predictive validity) (see also [AERA et al., 2014](#)).

As stated above, commitment is a necessary condition for reliable observation of problem-solving skills. We observed that most participants were highly motivated to take part in the research. Between sessions, the children often asked when they could participate again and showed clear excitement when it was their turn. While this enthusiasm may have been partly driven by the individual attention they received, we asked them after each session whether they had enjoyed the activity and what they liked or disliked about it. Almost without exception, they reported positive experiences, although this did not always align with our observations. Achieving a solution may have mitigated some frustration experienced during the problem-solving process.

Strong intercoder reliability and across task consistency were established. This indicates that PROSPER can reliably be used to observe problem-solving behaviours across different tasks. For this study, we only focused on intercoder reliability. In future research, it would be interesting to apply the observation scheme repeatedly to assess test-retest reliability.

#### 4.3. Limitations and future directions

Despite the strengths outlined above, this study is not without its limitations. First, the study is based on the analysis of 36 recordings in a sample of 12 preschoolers including boys and girls and children with diverse linguistic background who were offered three different engineering problems. Although we ensured the representativeness of the sample and aimed for maximum heterogeneity between participants, it would be beneficial to include more participants to further investigate the generalisability of the observation scheme. In addition, applying the observation scheme with other STEM-problems can provide more evidence on the external validity of the instrument.

Second, data collection was conducted by a researcher. Although the researcher was involved in classroom activities to familiarise with the children in the days before and during data collection, and the interactions closely mimicked authentic teacher-child dialogue, there remains a gap compared to a real classroom situation. Having data collected by the teachers in the children's own classroom would be even more authentic and could potentially elicit different behaviours. It is therefore desirable to involve teachers in data collection in follow-up studies.

Third, we decided to minimise guidance throughout the process to observe what preschoolers did on their own, while separating exploring and reflecting from the rest of the process. Closer involvement may be beneficial to better observe representing and planning behaviours.

Finally, the current observation scheme does not provide a systematic approach to translate observed behaviours into assessments of preschoolers' problem-solving competencies. Follow-up research could explore how the observation scheme can be used for assessment purposes.

## 5. Conclusion

This study contributes to the field of early childhood research and education by operationalising the PISA framework for use in preschool by assigning observable problem-solving behaviours to the different steps of the problem-solving process. It extends the PISA framework with four generic, overarching behaviours: commitment, perseverance, autonomy, and responsiveness. Although the present study focuses on preschool children aged 4-6 years solving engineering design problems, it is assumed that the observation instrument can be applied to other types of STEM problems and throughout early childhood (3-8 years).

This study is foundational for further research on problem solving in early childhood education by developing PROSPER, a comprehensive, valid, and reliable observation instrument that makes preschoolers' problem-solving behaviours visible in authentic contexts. Researchers can use the instrument to further investigate the development of preschooler's problem-solving competencies, to investigate the effectiveness of problem-solving interventions, and to investigate the relationship between problem solving and other constructs. However, given the complex nature of the problem-solving process and the comprehensiveness of the observation scheme, researchers should be aware that implementing this observation scheme is both challenging and time-consuming, necessitating training and substantial practice.

For educational practitioners, PROSPER provides a useful lens to observe problem solving as it occurs in the classroom. Focused observation can help preschool teachers in monitoring the development of preschoolers' problem-solving competencies, identify areas that require attention, and guide them in selecting or developing appropriate learning activities. Presenting children with challenging problems, observing their problem-solving behaviours, and engaging with the children during the problem-solving process provides researchers and educators valuable insights in children's problem-solving competencies. This way, the use of PROSPER in future (intervention) studies and educational practice can help to deepen insights into the development of children's problem-solving competencies and how to optimally stimulate this development, enabling future generations to meet the challenges of vital STEM careers.

### Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT, Copilot, Claude, DeepL and DeepL Write in order to check and improve writing and to find suitable labels for the codes. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Funding

This work was supported by the Research Foundation, Flanders (Fonds Wetenschappelijk Onderzoek, Vlaanderen) [Grant Number 1S03525N].

## Ethical approval

The research was approved by the Ethical Advisory Committee for Social and Human Sciences (EASHW) of the University of Antwerp on 26 September 2023 (file SHW\_2023\_243\_1). The parents of all the children involved gave written consent and allowed us to record the sessions.

## CRedit authorship contribution statement

**Joris Van Elsen:** Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Lisanne Buckers:** Validation, Formal analysis. **Charlotte Van Tricht:** Validation. **Joke Torbeyns:** Writing – review & editing, Supervision, Conceptualization. **Sven De Maeyer:** Writing – review & editing, Supervision, Conceptualization.

## Declaration of competing interest

The authors declare that they have no conflict of interest.

## Acknowledgements

The authors would like to thank the school team of Sint-Calasanz, Stabroek for their warm welcome and making a class available for the research. Special thanks to Anja Kenis for facilitating data collection. We would also like to thank Lieve Raeymaekers (teacher at Maria Boodschap, Antwerp), Ann Possemiers, Ann Verdonck, Evy Salsmans, and Magali Van den Bergh (teacher at Gibo Mariaburg) for processing the video recordings, and Ann Leysens (headmaster at Gibo Mariaburg) for making this possible. We are aware that processing recordings is time consuming, but involving experienced teachers was of great value for this study. Finally, we would like to thank all the children who solved our problems, as well as their parents for allowing us to record them on video.

## Supplementary materials

Supplementary materials associated with this article can be found on Open Science Framework (OSF) at [doi:10.17605/OSF.IO/K6X5R](https://doi.org/10.17605/OSF.IO/K6X5R).

## Appendix A. Some solutions for the three engineering design problems

Figs. A.1, A.2 and A.3



Fig. A.1. Volcano Bridge Problem.



Fig. A.2. Hook-a-Duck Problem.



Fig. A.3. The Storm Problem.

**Appendix B. The Problem-solving Observation Scheme for Playful Engineering and the Real-world (PROSPER)**

| Category      | Behaviour   | Description  | Examples of identified behaviours (“The child ...”)  |
|---------------|---|--|--|
| Exploring     | Observing   | Deliberately gathering information about the problem, materials, and possibilities by observation and interaction with the problem and environment | Takes a closer look at the set-up and the duck in bottle   |
|               | Naming  | Identifies materials by naming them  | Points to the materials as she names them  |
|               | Describing  | Specifies characteristics or functions of materials or tools   | Describes the volcano and what it is made of   |
|               | Comparing   | Examines or groups materials based on similarities or differences  | Compares the sheep as he holds them: “this one is a baby, and this one is bigger   |
|               | Associating   | Articulates links between objects or situations and other objects or situations  | Says: “What is this? The cow likes it. It is from the farm,” as he looks at the straw  |
|               | Manipulating  | Handles objects by touching them, moving them and/or demonstrating which actions can be performed with them  | Takes the ballpoint pen out of the toolbox, presses the button, finds that it does not work and says: “it’s off.”  |
|               | Searching   | Rummages through the materials or looks around in search for things that can potentially be used   | Rummages in the toolbox in search for what he can use  |
| Understanding | Asking information  | Asks questions about the available materials, the problem, or the criteria   | Takes a screw from the toolbox and asks: “What would this be? I don’t know what this is”.  |
|               | Listening   | Demonstrating understanding of the problem, its key elements, constraints and success criteria   | Listens and watches carefully at first, but wants to start before Mr Rabbit has finished explaining the problem and criteria.  |
|               | Confirming  | Pays attention to the problem description or clarifications of the goal and success criteria   | Nods to confirm when Mr Rabbit asks if she can make a bridge.  |
|               | Repeating   | Shows understanding by nodding or verbal expression  | Demonstrates what the kitten wants to do by going over the volcano with the kitten in the air.   |
| Complying     | Acts in accordance with the goal or success criteria  | Continues to place items on and against the volcano, even after repeated reminders.  |  |
| Representing  | Creating a mental, verbal, visual and/or physical representation of the problem and/or a goal situation                   |  |  |
| Formulating   | Thinking up   | Coming up with possible ways to solve the problem and evaluating alternatives  | Comes up with different structures: 2 rolls on top of each other, 2 rolls next to each other, a ramp, a staircase, a third pillar...                                   |
|               | Considering   | Articulates or shows a (new) idea or (other) possible solution   | Holds a [popsicle] stick to a pipe cleaner and says: “Then we can stick these sticks together like this. But I can’t paste this (...) because we don’t have any glue.” |
|               | Asking ideas  | Evaluates alternative options or substantiates the choice for a specific solution or strategy  | Turns to the supervisor and asks: “How did the others do it? Aren’t you allowed to tell? But I want you to tell me.”   |
|               | Asking ideas  | Aks for help in coming up with ideas   |  |
| Planning      | Preparing   | Determining (sub)goals, action plans, intermediate steps, and next steps   | Says: “Now I’m going to take this big stone,” while taking the biggest stones from the box and putting them next to it.  |
|               | Thinking ahead  | Prepares work by collecting and arranging materials, tidying up ...  | Says: “First I must make a wall. And then with a stick ...”  |
| Executing     | Doing   | Anticipates next step(s) or event(s)   |  |
|               | Doing   | Carrying out a plan, working towards a solution or trying to do so   | Works on through trial and error, gradually improving the construction.  |
|               | Refining  | Works towards a solution through physical actions  | Turns all cards white side up throughout the process.  |
| Monitoring    | Refining  | Finishes the result with care by adding extra elements or details  |  |
|               | Asking motor help   | Asks for concrete, motor assistance  | Fails to cut the tape and asks: “Can you hold this for a moment?” while handing the roll to the supervisor   |
| Monitoring    | Monitoring progress, identifying problems, interim check of success criteria, and adjusting plans or actions if necessary |  |  |

(continued on next page)

(continued)

| Category       | Behaviour           | Description   | Examples of identified behaviours ("The child ...")  |
|----------------|---------------------|---|--|
| Reflecting     | Testing             | Checks if the success criteria are being met  | Tests with hairdryer and says: "Here comes the storm again."   |
|                | Measuring           | Fits, measures, weighs, counts things to check whether they are the correct size, height, weight ...  | Measures the height of the stacked wooden discs against the stones and the sheep with his hand and says: "This must be as high as the stones (...). Must be above the sheep" |
|                | Determining         | Indicates whether something is working or not, or if there is a problem   | After accidentally dropping a spoon into the bottle, he says: "First I have to get the spoon out."   |
|                | Modifying           | Adjusts elements of the solution or own actions to improve it   | Reinforces the construction by adding a roll at each level   |
|                | Declaring           | Indicates that a solution has been reached  | Fishes the duckling out of the bottle, shouts "Yaa!!!" and immediately calls Mr Rabbit.  |
|                | Asking confirmation | Asks for approval or agreement from the supervisor  | Looks at the supervisor and asks: "Should it be like this?"  |
|                | Recalling           | Looking back at the problem and process and critically evaluating the outcome to learn from it  | Tells upon request what she had to do in the previous task, who she had to do it for, and names some materials she was allowed to use.                                       |
|                | Evaluating solution | Describes or demonstrates key elements from the previous assignment   | Spontaneously demonstrates to Mr Rabbit that the stable can withstand the storm by aiming the hairdryer at it.   |
|                | Evaluating process  | Describes or demonstrates the solution, demonstrates how it works, appraises the solution based on the criteria or suggests possible improvements | Recalls several failed trials and says the pencil was too short and the spoon had fallen into the bottle   |
|                | Generic             | Commitment  | Describes or demonstrates trials, concrete events, situations, difficulties or emotions during the problem-solving process   |
| Perseverance   |                     | <i>Overarching problem-solving behaviours</i><br>Is willing to solve the problem and is focused on the task                                       | Starts playing/throwing the kitten.  |
| Autonomy       |                     | Keeps going and does not give up when faced with setbacks or impasse  | Works patiently despite repeated collapses.  |
| Responsiveness |                     | Works independently with only limited need for support  | Regularly asks for help, then waits; does not come up with good plan of action himself   |
|                |                     | Adjusts actions based on tips and suggestions from the supervisor   | Does not initially follow the supervisor's suggestions. Follows only when the supervisor gives very specific instructions and helps physically.                              |

## Data availability

Data will be made available on request.

## References

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anggoro, F. K., Dubosarsky, M., & Kabourek, S. (2021). Developing an observation tool to measure preschool children's problem-solving skills. *Education Sciences*, 11(12), 779. <https://doi.org/10.3390/educsci11120779>
- Arik, M., & Topçu, M. S. (2022). Implementation of engineering design process in the K-12 science classrooms: trends and issues. *Research in Science Education*, 52(1), 21–43. <https://doi.org/10.1007/s11165-019-09912-x>
- Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, 4, 87–95. <https://doi.org/10.1007/s11409-009-9035-7>
- Bartholomew, S. R., Moon, C., Ruesch, E. Y., & Strimel, G. J. (2019). Kindergarten student's approaches to resolving open-ended design tasks. *Journal of Technology Education*, 30(2), 90–115. <https://doi.org/10.21061/jte.v30i2.a.6>
- Bat Or, M., Ben-Shoshan chen, M., & Shalev, O. (2022). Problem-solving expressions in preschool children's "Person Picking an Apple from a Tree" drawings and verbal narratives. *The Arts in Psychotherapy*, 81, Article 101964. <https://doi.org/10.1016/j.aip.2022.101964>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6(149). <https://doi.org/10.3389/fpubh.2018.00149>
- Bonner, S., Chen, P., Jones, K., & Milonovich, B. (2021). Formative assessment of computational thinking: cognitive and metacognitive processes. *Applied Measurement in Education*, 34(1), 27–45. <https://doi.org/10.1080/08957347.2020.1835912>
- Brassard, M. R., & Boehm, A. E. (2007). *Preschool assessment. Principles and practice*. Guilford.
- Bryce, D., & Whitebread, D. (2012). The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, 7(3), 197–217. <https://doi.org/10.1007/s11409-012-9091-2>
- Bustamante, A. S., Greenfield, D. B., & Nayfeld, I. (2018). Early childhood science and engineering: engaging platforms for fostering domain-general learning skills. *Education Sciences*, 8(3), 144. <https://www.mdpi.com/2227-7102/8/3/144>.
- Cartwright, K. (2023). Interpreting young children's multiplicative strategies through their drawn representations. *Mathematics Education Research Journal*. <https://doi.org/10.1007/s13394-023-00450-4>
- Csapó, B., & Funke, J. (2017). *The nature of problem solving: Using research to inspire 21st century learning*. OECD Publishing. <https://doi.org/10.1787/9789264273955-en>
- Cunningham, C. M., Lachapelle, C. P., & Davis, M. E. (2018). Engineering concepts, practices, and trajectories for early childhood education. In L. English, & T. Moore (Eds.), *Early engineering learning* (pp. 135–174). Singapore: Springer. [https://doi.org/10.1007/978-981-10-8621-2\\_8](https://doi.org/10.1007/978-981-10-8621-2_8).
- De Raadt, A., Warrens, M. J., Bosker, R. J., & Kiers, H. A. L. (2019). Kappa coefficients for missing data. *Educ Psychol Meas*, 79(3), 558–576. <https://doi.org/10.1177/0013164418823249>

- Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, 53(6), 1024–1037. <https://doi.org/10.1037/0022-3514.53.6.1024>
- DeLuca, C., Pyle, A., Valiquette, A., & LaPointe-McEwan, D. (2020). New directions for kindergarten education: embedding assessment in play-based learning. *The Elementary School Journal*, 120(3), 455–479. <https://doi.org/10.1086/707008>
- Dewey, J. (1910). *How we think*. D. C. Heath & CO.. <https://www.gutenberg.org/files/37423/37423-h/37423-h.htm>
- Dorie, B. L., Tranby, Z., Van Cleave, S. K., Cardella, M., & Svarovsky, G. N. (2013). Using puppets to elicit talk during interviews on engineering with young children. In P. University (Ed.), *School of Engineering Education Graduate Student Series*.
- Dubosarsky, M., John, M. S., Anggoro, F., Wunnava, S., & Celik, U. (2018). Seeds of STEM: the development of a problem-based STEM curriculum for early childhood classrooms. In L. English, & T. Moore (Eds.), *Early engineering learning* (pp. 249–269). Singapore: Springer. [https://doi.org/10.1007/978-981-10-8621-2\\_12](https://doi.org/10.1007/978-981-10-8621-2_12).
- English, L. D. (1992). *Children's use of domain-specific knowledge and domain-general strategies in novel problem solving*. <https://files.eric.ed.gov/fulltext/ED343952.pdf>.
- English, L. D. (2018). Early engineering: an introduction to young children's potential. In L. English, & T. Moore (Eds.), *Early engineering learning* (pp. 1–5). Singapore: Springer. [https://doi.org/10.1007/978-981-10-8621-2\\_1](https://doi.org/10.1007/978-981-10-8621-2_1).
- Epstein, A. S. (2003). How planning and reflection develop young children's thinking skills. *YC Young Children*, 58(5), 28–36. <http://www.jstor.org/stable/42728978>.
- Escolano-Pérez, E., Herrero-Nivela, M. L., & Anguera, M. T. (2019). Preschool metacognitive skill assessment in order to promote educational sensitive response from mixed-methods approach: complementarity of data analysis. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01298>
- European Commission. (2019). *Key competences for lifelong learning*. Publications Office of the European Union. <https://doi.org/10.2766/569540>
- Evans, N. S., Todaro, R. D., Schlesinger, M. A., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Examining the impact of children's exploration behaviors on creativity. *Journal of Experimental Child Psychology*, 207. <https://doi.org/10.1016/j.jecp.2021.105091>
- Fusaro, M., & Smith, M. C. (2018). Preschoolers' inquisitiveness and science-relevant problem solving. *Early Childhood Research Quarterly*, 42, 119–127. <https://doi.org/10.1016/j.ecresq.2017.09.002>
- Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: various coefficients of interrater reliability and agreement*. R package version 0.84.1. [R package]. <https://CRAN.R-project.org/package=irr>.
- Geist, E. (2025). Reinventing STEM. *Reinventing STEM in early childhood education: A constructivist approach* (pp. 1–31). <https://doi.org/10.4324/9781003488873-1>
- Gold, Z. S., Elicker, J., Evich, C. D., Mishra, A. A., Howe, N., & Weil, A. E. (2021). Engineering play with blocks as an informal learning context for executive function and planning. *Journal of Engineering Education*, 110(4), 803–818. <https://doi.org/10.1002/jee.20421>
- Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.
- Gunawan, I., Darhim, D., & Kusnandi, K. (2019). Exploration of the behavior of understanding mathematical concepts of junior high school students. *Journal of Physics: Education Series*, 1157(4), Article 042098. <https://doi.org/10.1088/1742-6596/1157/4/042098>
- Hayes, J. R. (1981). *The Complete Problem Solver*. Routledge. <https://doi.org/10.4324/9780203062715>
- Hudson, J. A., Sosa, B. B., & Shapiro, L. R. (1997). Scripts and plans: the development of preschool children's event knowledge and event planning. In S. L. Friedman, & E. K. Scholnick (Eds.), *The developmental psychology of planning. Why, how, and when do we plan?* (pp. 77–102). Lawrence Erlbaum Associates. [https://www.researchgate.net/publication/232426879\\_Scripts\\_and\\_plans.The\\_development\\_of\\_preschool\\_childrens\\_event\\_knowledge\\_and\\_event\\_planning](https://www.researchgate.net/publication/232426879_Scripts_and_plans.The_development_of_preschool_childrens_event_knowledge_and_event_planning).
- Isabelle, A. D., Russo, L., & Velazquez-Rojas, A. (2021). Using the engineering design process (EDP) to guide block play in the kindergarten classroom: exploring effects on learning outcomes. *International Journal of Play*, 10(1), 43–62. <https://doi.org/10.1080/21594937.2021.1878772>
- Jirout, J., & Klahr, D. (2020). Questions – and some answers – about young children's questions. *Journal of Cognition and Development*, 21(5), 729–753. <https://doi.org/10.1080/15248372.2020.1832492>
- Johnson, C. C., Mohr-Schroeder, M. J., Moore, T. J., & English, L. D. (2020). *Handbook of research on STEM education*. Routledge.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Jonassen, D. H. (2010). *Learning to solve problems. A handbook for designing problem-solving learning environments*. Routledge. <https://doi.org/10.4324/9780203847527>
- Kangas, K., Sormunen, K., & Korhonen, T. (2022). Creative learning with technologies in young students' steam education. In S. Papadakis, & M. Kalogiannakis (Eds.), *STEM, robotics, mobile apps in early childhood and primary education: Technology to promote teaching and learning* (pp. 157–179). Singapore: Springer Nature. [https://doi.org/10.1007/978-981-19-0568-1\\_9](https://doi.org/10.1007/978-981-19-0568-1_9)
- Lile Diamond, L. (2018). Problem solving in the early years. *Intervention in School and Clinic*, 53(4), 220–223. <https://doi.org/10.1177/1053451217712957>
- Liljedahl, P. (2016). Building thinking classrooms: conditions for problem-solving. In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Posing and solving mathematical problems: Advances and new perspectives* (pp. 361–386). Springer International Publishing. [https://doi.org/10.1007/978-3-319-28023-3\\_21](https://doi.org/10.1007/978-3-319-28023-3_21)
- Lippard, C. N., Lamm, M. H., & Riley, K. L. (2017). Engineering thinking in prekindergarten children: A systematic literature review. *Journal of Engineering Education*, 106(3), 454–474. <https://doi.org/10.1002/jee.20174>
- Lippard, C. N., Lamm, M. H., Tank, K. M., & Choi, J. Y. (2019). Pre-engineering thinking and the engineering habits of mind in preschool classroom. *Early Childhood Education Journal*, 47(2), 187–198. <https://doi.org/10.1007/s10643-018-0898-6>
- Luce, M. R., & Hsi, S. (2015). Science-relevant curiosity expression and interest in science: an exploratory study. *Science Education*, 99(1), 70–97. <https://doi.org/10.1002/sc.21144>
- MacDonald, B. L., Tofel-Grehl, C., & Searle, K. A. (2022). Play, problem-solving, stem conceptions, and efficacy in stem: an introduction to the stem in early childhood education special issue. *Education Sciences*, 12(5), 352. <https://www.mdpi.com/2227-7102/12/5/352>
- Malcok, B. A., & Ceylan, R. (2022). The effects of STEM activities on the problem-solving skills of 6-year-old preschool children. *European Early Childhood Education Research Journal*, 30(3), 423–436. <https://doi.org/10.1080/1350293x.2021.1965639>
- Marulis, L. M., & Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. *Metacognition and Learning*, 16(1), 207–231. <https://doi.org/10.1007/s11409-020-09244-6>
- Mayer, R. E. (1994). Problem solving. In M. W. Eysenck (Ed.), *The Blackwell dictionary of cognitive psychology* (pp. 284–288). Blackwell.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*, 22(3), 276–282.
- McLeod, B. D., Sutherland, K. S., Broda, M. D., Granger, K. L., Hollins, N., Frey, A. J., ... Dear, E. (2024). Initial development of the child responsiveness scale for early childhood settings. *Journal of Emotional and Behavioral Disorders*. <https://doi.org/10.1177/10634266241271392>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., ... de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Monteira, S. F., & Jiménez-Aleixandre, M. P. (2016). The practice of using evidence in kindergarten: The role of purposeful observation. *Journal of Research in Science Teaching*, 53(8), 1232–1258. <https://doi.org/10.1002/tea.21259>
- Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., Smith, K. A., & Stohlmann, M. S. (2014). A framework for quality k-12 engineering education: research and development. *Journal of Pre-College Engineering Education Research (J-PEER)*, 4(1). <https://doi.org/10.7771/2157-9288.1069>. Article 2.
- Moore, T. J., Johnson, A. C., & Glancy, A. W. (2020). STEM integration. A synthesis of conceptual frameworks and definitions. In C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & L. D. English (Eds.), *Handbook of Research on STEM Education* (pp. 3–16). Routledge.
- Moore, T. J., Tank, K. M., & English, L. (2018). Engineering in the early grades: harnessing children's natural ways of thinking. In L. English, & T. Moore (Eds.), *Early engineering learning* (pp. 9–18). Singapore: Springer. [https://doi.org/10.1007/978-981-10-8621-2\\_2](https://doi.org/10.1007/978-981-10-8621-2_2)
- NRICH. (2013, 05/09/2013). *Low threshold high ceiling - an introduction*. <https://nrich.maths.org/articles/low-threshold-high-ceiling-introduction>.
- OECD. (2004). *The PISA 2003 assessment framework: mathematics, reading, science and problem solving knowledge and skills*.
- OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. <https://doi.org/10.1787/9789264190511-en>.
- Polya, G. (1945). *How we solve it. A new aspect of mathematical method*. Princeton.

- Pouw, W. T. J. L., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, 26(1), 51–72. <https://doi.org/10.1007/s10648-014-9255-5>
- Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 3–30). Cambridge University Press. <https://doi.org/10.1017/CBO9780511615771.002>.
- Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., & Zu Belzen, A. U. (2020). A framework to foster problem-solving in STEM and computing education. *Research in Science & Technological Education*, 38(1), 105–130. <https://doi.org/10.1080/02635143.2019.1600490>
- Purzer, Ş., & Douglas, K. A. (2018). Assessing early engineering thinking and design competencies in the classroom. In L. English, & T. Moore (Eds.), *Early engineering learning* (pp. 113–132). Singapore: Springer. [https://doi.org/10.1007/978-981-10-8621-2\\_7](https://doi.org/10.1007/978-981-10-8621-2_7).
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (4.3.2) <https://www.R-project.org/>.
- Reed, S. K. (2016). The structure of ill-structured (and well-structured) problems revisited. *Educational Psychology Review*, 28(4), 691–716. <https://doi.org/10.1007/s10648-015-9343-1>
- Rhodes, K. T., Richland, L. E., & Alcalá, L. (2024). Problem solving is embedded in context... so how do we measure it? *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1380178>
- Rogers, J. (2004). Autonomy and mathematical problem-solving: the early years. *Education 3-13*, 32(3), 24–31. <https://doi.org/10.1080/03004270485200291>
- Rott, B., Specht, B., & Knipping, C. (2021). A descriptive phase model of problem-solving processes. *ZDM – Mathematics Education*, 53(4), 737–752. <https://doi.org/10.1007/s11858-021-01244-3>
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2 ed.). Sage.
- Salminen-Saari, J. F. A., Garcia Moreno-Esteva, E., Haataja, E., Toivanen, M., Hannula, M. S., & Laine, A. (2021). Phases of collaborative mathematical problem solving and joint attention: a case study utilizing mobile gaze tracking. *ZDM – Mathematics Education*, 53(4), 771–784. <https://doi.org/10.1007/s11858-021-01280-z>
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Shechter, T., Eden, S., & Spektor-Levy, O. (2021). Preschoolers' nascent engineering thinking during a construction task. *Journal of Cognitive Education and Psychology*, 20(2), 83–111. <https://doi.org/10.1891/JCEP-D-20-00010>
- Shiakalli, M. A., & Zacharos, K. (2012). The contribution of external representations in pre-school mathematical problem solving. *International Journal of Early Years Education*, 20(4), 315–331. <https://doi.org/10.1080/09669760.2012.714992>
- Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4(3), 181–201. [https://doi.org/10.1016/0004-3702\(73\)90011-8](https://doi.org/10.1016/0004-3702(73)90011-8)
- Skinner, B. F. (1953). *Science and human behavior*. Free Press.
- Stoll, J., Hamilton, A., Oxley, E., Eastman, A. M., & Brent, R. (2012). Young thinkers in motion: problem solving and physics in preschool. *YC Young Children*, 67(2), 20–26. <http://www.jstor.org/stable/42731148>.
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., ... Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, 27(5), 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Thornberg, R., & Charmaz, K. (2014). Grounded theory and theoretical coding. In U. Flick (Ed.), *Grounded theory and theoretical coding* (pp. 153–169). SAGE Publications Ltd. <https://doi.org/10.4135/9781446282243.n11>.
- Thunder, K., Almrode, J., & Hattie, J. (2022). *Visible learning in early childhood*.
- Tytler, R. (2020). Stem education for the twenty-first century. In J. Anderson, & Y. Li (Eds.), *Integrated approaches to stem education: An international perspective* (pp. 21–43). Springer International Publishing. [https://doi.org/10.1007/978-3-030-52229-2\\_3](https://doi.org/10.1007/978-3-030-52229-2_3).
- Uher, J. (2016). What is behaviour? And (when) is language behaviour? A metatheoretical definition. *Journal for the Theory of Social Behaviour*, 46(4), 475–501. <https://doi.org/10.1111/jtsb.12104>
- UNESCO Institute for Statistics. (2012). *International standard classification of education (ISCED) 2011*. <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>.
- Vessonen, T., Hellstrand, H., Aunio, P., & Laine, A. (2023). Individual differences in mathematical problem-solving skills among 3- to 5-year-old preschoolers. *International Journal of Early Childhood*. <https://doi.org/10.1007/s13158-023-00361-2>
- Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Harvard University Press.
- Warneken, F., Steinwender, J., Hamann, K., & Tomasello, M. (2014). Young children's planning in a collaborative problem-solving task. *Cognitive Development*, 31, 48–58. <https://doi.org/10.1016/j.cogdev.2014.02.003>
- Weiser, L. E. (2022). Young children's free play in nature: an essential foundation for stem learning in germany. In S. D. Tunnicliffe, & T. J. Kennedy (Eds.), *Play and STEM Education in the Early Years: International Policies and Practices* (pp. 85–103). Springer International Publishing. [https://doi.org/10.1007/978-3-030-99830-1\\_4](https://doi.org/10.1007/978-3-030-99830-1_4).
- Wheatley, C. L., & Wheatley, G. H. (1984). Problem solving in the primary grades. *The Arithmetic Teacher*, 31(8), 22–25. <http://www.jstor.org/stable/41192375>.
- Wong, M., Choi, K., Barak, L., Lapidow, E., Austin, J., Shafto, P., & Bonawitz, E. (2024). Young children's directed question asking in preschool classrooms. *Behavioral Sciences*, 14(9), 754. <https://www.mdpi.com/2076-328X/14/9/754>.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Yuan, Y., Zeng, W., Kloos, H., Brown, R., & Carr, V. (2024). Preschool engineering play on nature playscapes. *Early Childhood Education Journal*. <https://doi.org/10.1007/s10643-024-01743-4>
- Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, 1(2), 198–226. <https://doi.org/10.1037/1089-2680.1.2.198>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation. Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). Routledge.