Check for updates

Received: 10 March 2025 | Accepted: 1 June 2026

DOI: 10.1002/rev3.70181

REVIEW AND SYNTHESIS

Review of Education BERA

Methods for assessing problem-solving competencies in preschoolers: A scoping review

Joris Van Elsen[1] | Sven De Maeyer[1] | Joke Torbeyns[2]

[1]Department of Training and Education Sciences, University of Antwerp, Antwerp, Belgium

[2]Center for Instructional Psychology and Technology, KU Leuven, Leuven, Belgium

Correspondence
Joris Van Elsen, University of Antwerp, Sint Jacobsstraat 2 – Room Z.408, 2000 Antwerp, Belgium.
Email: joris.vanelsen@uantwerpen.be

Funding information
Fonds Wetenschappelijk Onderzoek, Grant/Award Number: 1S03525N

Abstract

Problem-solving is an important 21st-century skill and an indispensable part of early STEM (science, technology, engineering and mathematics) education. To effectively study the development of children's problem-solving competencies and to evaluate the effectiveness of classroom interventions and programmes, researchers and educators need reliable and valid assessment instruments. This scoping review aimed to identify and discuss methods that are developed, studied or used in scientific research to assess preschoolers' competencies in solving ill-structured problems in and across the four STEM disciplines. A systematic search of Web of Science, Scopus and ERIC identified 24 studies published between 1974 and 2024. These studies reported 18 unique methods for assessing problem-solving competencies in preschool children aged 3–6 years. Results indicate that most assessment methods rely on observation of hands-on tasks or oral examination, typically focussing on a single step of the problem-solving process within one STEM discipline. Observing children solve engineering design problems and engaging in dialogue with them appears appropriate for integrated problem-solving assessment. Only one instrument was found for which the psychometric properties were adequately examined. Further research is needed to develop reliable and valid assessment methods to provide researchers and educators with age-appropriate methods to investigate and monitor preschoolers' problem-solving competencies.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

KEYWORDS
early childhood education, educational assessment, literature review, problem solving, STEM

Context and implications

Rationale for this study: Identifying characteristics of reliable and valid methods for assessing preschoolers' problem-solving competencies will help researchers and practitioners to study and support their development.
Why the new findings matter: The findings offer valuable insights into the current methods used to assess preschoolers' problem-solving competencies in scientific research, and draw attention to issues that need to be considered when selecting, developing or implementing assessment methods.
Implications for researchers, practitioners: From this scoping review, we conclude that there is still a need for reliable and valid, cross-disciplinary assessment methods that cover the whole problem-solving process and can be administered multiple times. Most of the methods identified involve oral examination of depicted problems or observation of hands-on assignments. Observing children as they solve engineering design problems and engaging in dialogue with them offers an integrated and promising approach for assessing and monitoring the development of preschoolers' problem-solving competencies.

INTRODUCTION

Problem-solving is a key 21st-century competency that is required for lifelong learning and empowers individuals to navigate our increasingly complex society (European Commission, 2019; Suto, 2013). Early childhood offers a critical window for developing problem-solving competencies through pedagogical activities (Burns et al., 2025; Lile Diamond, 2018; Malcok & Ceylan, 2022; O'Reilly et al., 2022; Rogers, 2004). Because developing problem-solving competencies is a long-term process spanning years or even decades (Molnár et al., 2013), stimulating these competencies from preschool (ages 3–6) onwards is strongly recommended (Fusaro & Smith, 2018; Lopes et al., 2017). Moreover, problem-solving plays a central role in science, technology, engineering and mathematics (STEM, Priemer et al., 2020), and integrating STEM-related problem solving into early childhood education provides ample opportunities to develop crucial problem-solving skills (MacDonald et al., 2022).

Researchers and educators need reliable and valid assessment methods to investigate, monitor and support the development of children's problem-solving competencies and to evaluate the effectiveness of classroom interventions and programmes (Snow & Van Hemel, 2008). Assessment refers to the process of collecting, interpreting and using evidence about student learning (Harlen, 2007; McMillan, 2013) and includes different kinds of measurements and testing (AERA et al., 2014; Popham, 2017). For researchers and educators, it is important to understand how problem-solving competencies are being assessed in preschoolers, including the methods used, their characteristics, their scope, the procedures and their psychometric properties.

Researchers and educators can only rely on assessment outcomes if the methods used are thoughtfully developed, considering crucial elements related to reliability and validity.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In their review on reporting practices related to reliability and validity of early childhood assessment, Barghaus et al. (2022) found that only a small fraction of assessment developers report on all required issues concerning reliability and validity as proposed in the *Standards for Educational and Psychological Testing* (AERA et al., 2014). In their systematic review and meta-analysis on approaches to teaching problem-solving skills in early childhood education and care settings, Burns et al. (2025) emphasise the importance of comprehensive, reliable and valid instruments to assess young children's problem-solving skills. However, it remains unclear to what extent reliable and valid instruments were used in the included studies.

The present scoping review aimed to provide an overview of methods developed, studied or used in early childhood education research to assess problem-solving competencies in preschool children (3–6 years). More specifically, this study aimed to identify and analyse the characteristics, scope, procedures and psychometric properties of these methods.

Before describing the methodology and results, we elaborate on the conceptual definition of problem solving in the context of STEM and the steps of the problem-solving process. We conclude with discussing the implications of our findings for research and practice.

Problem-solving in STEM

Problem-solving is the process of dealing with a problem where no obvious solution method is available (OECD, 2004). It encompasses a wide range of cognitive and non-cognitive competencies, such as information processing, representation, reasoning, self-regulation, planning and decision-making (Greiff et al., 2014). Given the diversity of potential problems and the observation that a routine job for one person may be an impossible problem for another person (Bransford & Barry, 1993), a situation or assignment is considered a problem when the subject is able to understand it, but currently lacks a procedure to solve it (Brownell, 1942).

Problem-solving is an indispensable part of STEM education (Priemer et al., 2020). STEM refers to educational activities within science, technology, engineering and mathematics (Bryan & Guzey, 2020). According to the International Technology and Engineering Educators Association (ITEEA, 2020), science can be understood as ‘Investigating and understanding the phenomena in the natural world’ (p. 159). Technology involves altering the environment through the creation of tools, systems or processes designed to fulfil human needs and desires. Engineering refers to the application of scientific knowledge and mathematical reasoning to design and refine technological solutions that meet specific criteria under constraints. Mathematics is thinking and reasoning about patterns, order, measurement, properties, quantities, numbers, operations and geometry (Milburn et al., 2019). STEM education can be approached in various ways, ranging from teaching discipline-specific knowledge and skills in isolation to integrating and applying them across multiple disciplines (English, 2016).

Young children exhibit a natural curiosity about STEM and already possess foundational STEM skills and emerging understanding of STEM concepts from an early age (Clements & Sarama, 2016). In preschool, STEM is often embedded in play, everyday experiences and enquiry- or problem-based activities (MacDonald et al., 2022; Tunnicliffe, 2022). Young children are encouraged to explore scientific phenomena such as light and shadow (Wang & Chen, 2025), explore technologies by examining artefacts, crafting and creating (Sundqvist & Nilsson, 2018), develop engineering concepts and practices through designing, building and constructing (Lippard et al., 2018) and learn mathematical concepts such as quantities, patterns, symmetry and dimensions through playful activities such as block play (Qiao & Huang, 2023).

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Authentic STEM problems tend to be ill-structured (Nadelson & Seifert, 2017). Ill-structured problems are 'fuzzy' problems that occur in everyday life (Simon, 1978). They are open-ended and allow for multiple solution paths, as there are no clearly defined rules or preferred algorithms to solve them (Frensch & Funke, 1995). Well-structured problems, on the other hand, typically have one or a limited number of correct solutions and can be solved by applying certain procedures or algorithms (Jonassen, 1997; Simon, 1973). Solving ill-structured problems involves different cognitive processes than solving well-structured ones and requires distinct learning environments (Jonassen, 2010). Therefore, findings derived from solving well-structured problems such as Tower of Hanoi-style puzzles or routine mathematical (word) problems cannot simply be generalised to solving ill-structured problems (Funke & Frensch, 2007).

Steps of the problem-solving process

Different problems require different solution methods and strategies. Although most researchers emphasise that problem-solving is not necessary a linear process, many scholars (e.g. Dewey, 1910; Polya, 1945; Schoenfeld, 1985) have decomposed the process to operationalise problem-solving in measurable and teachable phases or steps (Rott et al., 2021). The PISA Problem-Solving Framework builds on Polya's four-phase model. Although originally designed to assess the problem-solving skills of 15-year-olds, it is equally applicable to younger children. Its emphasises on exploratory and creative processes is particularly important for fostering effective problem-solving in early childhood (Evans et al., 2021). Rather than relying on domain-specific knowledge and skills, the framework focusses on engagement with authentic, real-world problems that align well with the principles of early STEM education (Moore et al., 2020).

The PISA framework divides the problem-solving process into eight sub-processes: Exploring, Understanding, Representing, Formulating, Planning, Executing, Monitoring and Reflecting (OECD, 2013). Exploring is conceptualised as observing and interacting with the problem, searching for information and finding limitation and obstacles. Understanding involves understanding given and obtained information and demonstrating understanding of relevant concepts. Representing refers to building a mental representation and constructing tabular, graphical, symbolic or verbal representations. Formulating is defined as generating hypotheses by identifying relevant elements of the problem and their interrelationships and critically evaluating information. Planning relates to setting and clarifying (sub-)goals and devising a plan or strategy to reach that goal. Executing refers to carrying out the plan. Monitoring refers to monitoring progress towards the goal, checking (intermediate) results, identifying unexpected events and adjusting as required. Reflecting, the final step in the problem-solving process, involves critically evaluating solutions from different perspectives, identifying additional information needs and communicating progress.

The present study

To monitor development of children's problem-solving competencies and to evaluate the effectiveness of classroom interventions and programmes, researchers and educators require reliable and valid assessment instruments (Burns et al., 2025). A comprehensive overview of available instruments, detailing their key characteristics, scope and psychometric properties, would provide essential guidance for researchers and educators in selecting, developing or implementing appropriate assessment methods.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Two recent review studies aimed to identify assessment methods to enhance students' problem-solving skills but did not address preschool education (Ukobizaba et al., 2021; Wicaksono & Korom, 2022). The same applies for recent review studies on assessment of related topics such as computational thinking (Cutumisu et al., 2019; Lu et al., 2022; Tang et al., 2020; Varghese & Renumol, 2021) and creative thinking (Bolden et al., 2020; Said-Metwaly et al., 2017; Suherman & Vidákovich, 2022). Systematic reviews examining assessment methods for preschool children address related concepts such as executive functions (Mehsen et al., 2022; Silva et al., 2022) or self-regulation (Philpott-Robinson et al., 2023), but not problem-solving. The recent systematic review and meta-analysis of Burns et al. (2025) addresses problem-solving in early childhood education (including preschool), but only pays limited attention to assessment. To the best of our knowledge, no comprehensive overview of methods developed, studied or used in scientific research to assess problem-solving competencies in preschoolers currently exists.

This scoping review aimed to fill this gap in scientific research by providing an overview of the methods developed, studied or used to assess the problem-solving competencies of preschool children (3–6 years), and to map the characteristics, scope, procedures and psychometric properties of these methods. A scoping review is an appropriate approach to map how a phenomenon is studied across research with varying methodological designs (Aromataris & Munn, 2020; Munn et al., 2018). Unlike systematic reviews, scoping reviews do not aim to answer a single, narrowly defined question. Instead, they aim to identify the characteristics of a particular phenomenon or to provide an overview of the existing evidence (Munn et al., 2018).

Synthesising the characteristics of existing assessment methods offers valuable insights into how problem-solving competencies are currently assessed in research and can support researchers and educators in selecting, developing or implementing appropriate assessment methods. Mapping the scope of these methods helps identify which aspects of the problem-solving process and STEM disciplines are being addressed, while also revealing gaps that warrant further investigation. Additionally, examining the procedures and psychometric properties of the methods provides important information about their usability, reliability and validity. The following research questions guide this study:

RQ 1. What are the characteristics of the methods developed, studied or used in scientific literature to assess problem-solving competencies in preschoolers?

RQ 2. Which steps of the problem-solving process and which STEM-disciplines are being assessed with these methods?

RQ 3. What are the procedures and psychometric properties of these methods?

METHOD

Design

The study was conducted according to the scoping review guidelines framework of the Joanna Briggs Institute (JBI, Aromataris & Munn, 2020). The scoping review methodology was chosen for a number of reasons. First, this study aimed to explore the field of research concerning methods for assessing problem-solving competencies in preschoolers, which is in line with the exploratory nature of scoping reviews (Christou et al., 2025; Munn et al., 2018; Tricco et al., 2018). Second, the scoping review approach provides an inclusive framework for bringing together diverse literature, including quantitative, qualitative and mixed-method studies (Aromataris & Munn, 2020). This approach enables inclusion of development studies, validation studies and intervention studies. Third, scoping reviews are particularly well-suited to identify characteristics of a phenomenon and how it is being investigated (Munn et al., 2018). This aligns with our aim to map and analyse the characteristics and psychometric properties of the methods that have been developed, studied or used in scientific research to assess problem-solving competencies in preschoolers. Finally, scoping reviews allow to identify gaps in scientific knowledge and can serve as a foundation for future systematic reviews (Munn et al., 2018). By mapping assessment methods developed, studied or used in scientific research, this study provides an overview of the current state of research on assessing problem-solving competencies in preschoolers and provides direction for future research into reliable and valid assessment methods.

Reporting is based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) statement (Tricco et al., 2018). A review protocol was pre-registered on Open Science Framework (OSF, https://doi.org/10.17605/OSF.IO/QPJAT). The protocol outlines the objectives, search and selection procedures, data extraction and analysis. Any deviations from the protocol are addressed in the Discussion section.

Search strategy

A three-stage search strategy was applied to identify relevant sources (Aromataris & Munn, 2020). First, a search string was derived from the research questions. A preliminary search on Web of Science was conducted and keywords of relevant studies were screened to identify additional search terms. In consultation with a university librarian, the search string was refined, and three databases were selected: Web of Science, Scopus and Eric. Second, a main search on title, abstract and keywords was conducted across all databases, using the following search string[i]:

```
(TITLE-ABS-KEY (problem* W/5 solv*) AND TITLE-ABS-KEY (preschool* OR kindergarten* OR "pre-k" OR "early educat*" OR "early school*" OR ecec OR "early childhood educat*" OR "nursery school") AND TITLE-ABS-KEY(assess* OR monitor* OR evaluat* OR testing* OR measur*)) AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (DOCTYPE, "ar")).
```

Finally, reference lists of selected reports were screened for additional sources (backward search) and sources that refer to the identified reports were traced through cited reference searching in Web of Science and Scopus (forward search). All identified sources were collated in EndNote (version 20.1) and duplicates were removed.

Eligibility criteria

Only peer-reviewed journal articles presenting empirical quantitative, qualitative and mixed method studies were considered. Restricting the sources to peer-reviewed articles was intended to ensure that the methods are grounded in scientific evidence. Only sources written in English were considered. Inclusion criteria were conceived according to the PCC framework (Participants/Population, Concept and Context) (Aromataris & Munn, 2020). We included studies that focus on children enrolled in preschool programmes with an intentional education component classified under ISCED 02: Early childhood education from age 3 to the start of primary education (UNESCO Institute for Statistics, 2012). Specifically, studies were retained if the data were collected in an institution described as 'preschool' or 'kindergarten', or if the data were collected in a school or classroom and at

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

least some of the participants were in the 3–6 age group. To be eligible, the studies must include problem-solving assessment that focus on ill-structured STEM problems. Because structuredness of a problem is a multidimensional continuum (Jonassen, 2010), studies were retained if at least one problem presented to the participants was non-routine and provided opportunities for multiple solutions or multiple solution paths. Studies that focus only on solving well-structured problems or problems with only one correct solution such as simple tasks (e.g. conservation tasks, balance beam tasks), prescribed operations (e.g. mathematical operations, coding problems) or puzzles (e.g. Tower of Hanoi, labyrinths) were excluded. Since this review focusses on STEM problems, only studies in which the problem-solving assessment relates to at least one of the STEM disciplines were considered eligible, thereby excluding social problems (e.g. interpersonal problems, conflict handling), and reasoning problems (e.g. problems related to causality, analogy, syllogisms). Finally, only studies that include educational assessment administered in preschool were eligible. Clinical studies, labo-experiments conducted outside the school and screenings for disabilities were excluded. Consequently, we did not include studies that make use of tests such as the Raven's Coloured Progressive Matrices (RCPM), Ages & Stages Questionnaires® (ASQ) or the Dimensional Change Card Sort (DCCS) for problem-solving assessment since they only pay limited attention to problem-solving and do not focus on STEM.

Study selection

AI-assisted screening and assessment of risk of bias by automation

Titles and abstracts were screened for eligibility by the first author using ASReview (version 1.0), a researcher-in-the-loop active-learning application that helps prioritising relevant studies (van de Schoot et al., 2021). ASReview is an open-source application that makes use of artificial intelligence to clusters and prioritises the most relevant sources, thereby facilitating the screening process and significantly reducing the time required for manually screening titles and abstracts.

The default Naïve Bayes model (TF-IDF, Naive Bayes, Maximum, Dynamic resampling) was used for screening. The model was trained on two previously identified relevant studies meeting the inclusion criteria and 10 randomly selected studies of which one was marked relevant and nine were rejected. During the screening process, all records labelled irrelevant were tagged with a reason why they were excluded. Records not presented by the algorithm were considered not eligible.

Two mutual independent stopping criteria were applied: at least 33% of the total data set has been screened (van de Schoot et al., 2021) and 100 consecutive records were labelled irrelevant (Wagner et al., 2022). The second criterion is more stringent than the recommendation of Ros et al. (2017) to stop after 50 consecutive rejections.

Risk of bias through automation was assessed by having the second and third authors independently screen a random sample of 10% of the records ($n = 250$). The results of the screening process by the first author were compared with those of the second and third authors. Percentage of agreement between raters was 93%. This aligns with the expected error rates for human reviewers as reported by Wang et al. (2020). All discrepancies were analysed and discussed within the research team.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Screening title and abstract in two rounds

Title and abstract screening was conducted in two rounds. In scoping reviews, an iterative team-based approach is recommended (Levac et al., 2010). During the first round of screening, we noticed that the concept problem-solving is used as an umbrella term to refer to resolving a wide range of tasks. We therefore decided to further specify the inclusion criteria and align them more closely with the conceptualisation of problem-solving described in the theoretical framework of this study. Based on a list of all encountered problem types that was composed by the first author during screening, the inclusion criteria were further specified. Studies focussing on well-structured problems, problems with only one correct answer, routine tasks, operations or screenings were excluded. For convenience, the eligibility criteria presented in the previous section are the final criteria.

For the second round of screening, six studies identified as relevant in the previous round and five randomly selected studies labelled as irrelevant were used as prior knowledge to train the AI model in ASReview. Sources marked as relevant in the second round were compared with those from the first round to verify that no relevant sources from the initial round were omitted. Once again, 10% randomly selected records were screened manually by the second and third author. Percentage of agreement between raters was 91%, which is again consistent with expected error rates in manual screening (Wang et al., 2020). Discrepancies were analysed and discussed. The differences were due to unclear or missing information in the abstract and all differently rated sources were rejected after discussion within the research team or full-text screening.

Full text screening

Full texts of eligible sources were retrieved through EndNote, databases, journal webpages and ResearchGate. If a full text was not readily accessible, authors were contacted through ResearchGate, email or LinkedIn. During full-text screening, assessment methods were scrutinised. If the study failed to provide sufficient information about the assessment method, additional information was looked up through the provided references and authors of the sources and instruments were contacted. All excluded studies were coded by the main reason of exclusion and doubtful cases were discussed with the research team.

Data extraction

A digital data extraction form was drafted in Qualtrics, pilot-tested and refined after discussion with the research team. To avoid errors, charting was performed twice. During the second round, relevant data were charted in pre-structured tables (MS Excel). Differences between the two rounds were double checked.

From the selected sources, information was extracted about the study characteristics (year of publication, educational system, study design, sample size, participants' age, research questions, theoretical framework), characteristics of the assessment method (description of the assessment, assessment type, assessment materials, objectives of the assessment, data types and scoring), scope of the problem-solving assessment (addressed problem-solving steps and STEM disciplines), details about the assessment procedures (administrator, location, duration and whether the session was done individual, in pairs or with the entire classroom) and psychometric properties of the included assessment instruments (types of reliability and validity, metrics). If the scope of the assessment method was unclear,

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

authors were contacted to obtain the relevant items or scale. All data on the sources and assessment methods are presented in Tables 1–5.

Synthesis

For each included assessment method, a short description was written by the first author in a standardised format. Frequency tables were generated to provide a summary of assessment types, materials used, data collection methods and outcomes, highlighting prevalence of the assessment characteristics. Typical and atypical methods were described qualitatively and reported in more depth. All included assessment methods were coded regarding to the problem-solving steps being assessed and the STEM disciplines being addressed. Where authors indicated that the method used addressed integrated STEM, all four STEM disciplines were checked. Frequency tables were created and prevalence of the steps and disciplines was analysed. The reported statistics on the psychometric properties of the assessment methods were collated in an Excel file and categorised by type of evidence.

RESULTS

Literature search

The main database search was conducted on 21 September 2022 and resulted in 2979 records. Figure 1 provides an overview of the selection process according to the PRISMA 2020 statement (Page et al., 2021). After deduplication ($n=521$), 2458 records were retained for title and abstract screening. The stopping rules were met after screening 1389 records of which 213 were marked relevant. The 1069 records that were not presented by the algorithm were labelled irrelevant.

Fourteen mainly older articles could not be retrieved. After full-text screening, 17 reports were included. Four additional sources were included through backward search. Forward search yielded no new records. On 2 October 2024, an additional search was conducted to retrieve sources that were published during the course of the study. This updated literature search resulted in 349 new reports of which three met the inclusion criteria. The 24 reports included in the final selection contain information about 18 different assessment methods.

Study characteristics

Study characteristics of the 24 studies included in this review are presented in Table 1. The studies were published between 1978 and 2024 with more than half of the studies published in the last 10 years ($n=16$). Studies were conducted in nine different countries. The Middle East is best represented in the selected studies ($n=10$), followed by North America ($n=7$), and Europe ($n=5$). Three studies were conducted in East Asia. There were no studies from Central and South America, South and Southeast Asia, Africa or Australia. Different designs were used in the studies, with most studies being (quasi-)experimental ($n=9$), correlational ($n=4$) or descriptive ($n=5$). The age of the children that participated in the studies ranged from 3 to 7 years. The study of German (2000) involved children aged 5, 6 and 7 of whom only the 5- and 6-year-olds are considered preschoolers. A total of 2219 preschoolers participated in the studies with an average of 92 (*min* = 5, *max* = 328) preschoolers per study.

The included studies report on 18 different methods to assess problem-solving competencies in preschool. Four methods were used in two or three studies. In one study, two

TABLE 1 Characteristics of included studies.

Study	Country	Design	Participants' age (years)	Sample size (preschoolers)	Origins of the assessment method
Alexander et al. (1994)	USA	Descriptive	NA	61	Ad hoc, based on Torrance (1974)
Anggoro et al. (2021)	USA	Test development	3–5	238	Ad hoc
Annevirta and Vauras (2006)	Finland	Longitudinal	6	43	Ad hoc
Bartholomew et al. (2019)	USA	Descriptive	5–6	55	Ad hoc
Cakir et al. (2021)	Türkiye	Experimental	5	40	Oğuz and Köksal- Akyol (2015)[a]
Demirel and Deretarla Gul (2021)	Türkiye	Quasi-experimental	4-5	32	Oğuz and Köksal- Akyol (2015)[a]
Fusaro and Smith (2018)	USA	Correlation	4–5	24	Ad hoc
Gardiner et al. (2012)	USA	Experimental	2–3	46	Ad hoc
German and Defeyter (2000)	UK	Experimental	5–7	60	Ad hoc, based on Duncker (1945)
Guevara et al. (2020)	Netherlands	Repeated measures	4–6	8	Ad hoc
Hong and Milgram (1991)	USA/Israel	Descriptive	4-6	48	Ad hoc and Ray (1955)
Leikin (2013)	Israel	Descriptive	3–5	37	Ad hoc, based on Leikin (2009) and Torrance (1974); Tsamir et al. (2010)
Leikin and Tovli (2014)	Israel	Correlation	5–6	31	Leikin (2013), based on Torrance (1974); Tsamir et al. (2010)
Lin et al. (2021)	China	Quasi-experimental	4-5	122	Fusaro and Smith (2018)
Lin et al. (2023)	China	Correlation	5	226	Fusaro and Smith (2018)
Nam et al. (2019)	South Korea	Quasi-experimental	5–6	53	Adapted from Ward (1993)
Parsonson and Baer (1978)	USA	Multiple-baseline	3-6	5	Ad hoc
Shechter et al. (2021)	Israel	Correlation	4–6	228	Ad hoc

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Study	Country	Design	Participants' age (years)	Sample size (preschoolers)	Origins of the assessment method
Tsamir et al. (2010)	Israel	Descriptive	5–6	163	Ad hoc
Unal and Aral (2014a)	Türkiye	Test development	5	174	Ad hoc
Unal and Aral (2014b)	Türkiye	Quasi-experimental	5	42	Unal and Aral (2014a)
Unal and Saglam (2018)	Türkiye	Quasi-experimental	5	25	Unal and Aral (2014a)
Van Elsen et al. (2024)	Belgium (Flanders)	Quasi-experimental	5–6	130	Ad hoc, based on Leikin (2013)
Vessonen et al. (2024)	Finland	Descriptive	3–5	328	Ad hoc

Note: NA = Data not specified in report.
[a]Study written in Turkish.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

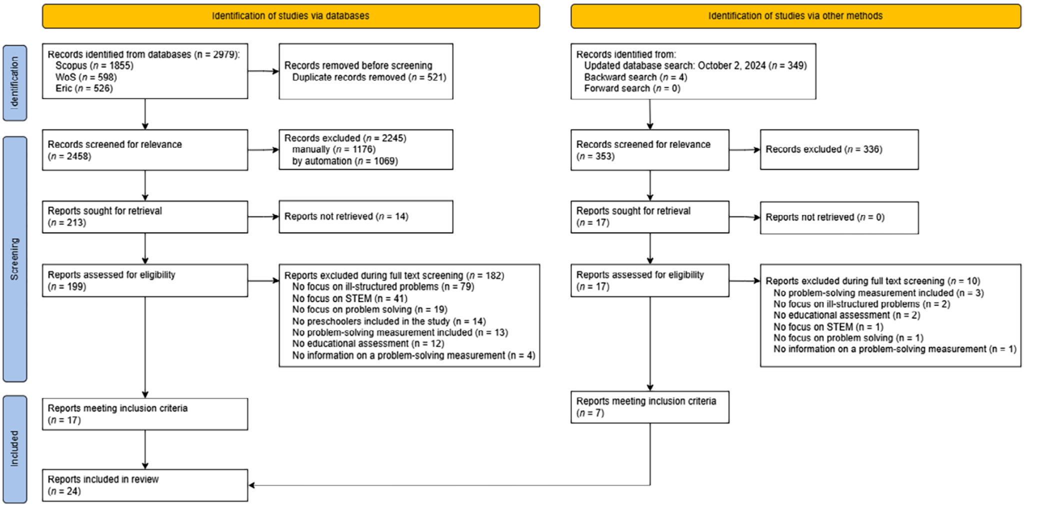

FIGURE 1 PRISMA 2020 flow diagram (Page et al., 2021).

methods were used to assess problem-solving competencies (Leikin & Tovli, 2014). Most studies include an assessment method purposefully designed to meet the needs of the study ($n = 14$). Five of these methods were adaptations of, or based on, existing methods. Ten studies made use of an existing method and two studies aimed at developing a new assessment method or instrument.

Characteristics of the identified assessment methods

To answer the first research question, we shift the unit of analysis from the individual study to the assessment methods that are used in these studies. Table 2 provides a short description of the assessment methods and their characteristics. Nearly all assessment methods make use of hands-on assignments ($n = 10$) or are based on oral examination ($n = 7$). The Design Portfolio's (Bartholomew et al., 2019) apply portfolio assessment.

Hands-on assignments

Generally, hands-on assignments make use of commonly available crafts materials (e.g. cardboard, tape, glue, bottle caps, pipe cleaners) ($n = 5$). Other materials used are building bricks (e.g. Lego), toys, tools and consumer goods. Two methods make use of a specially designed apparatus. Hands-on assignments are most often accompanied by observation for data collection. These observations were conducted directly ($n = 5$), via video recordings ($n = 4$) or combined both approaches ($n = 1$). The Stringent Solution-Standard Measures (Hong & Milgram, 1991) makes use of oral examination to collect data during the hands-on assignment. Observations of hands-on assignments typically result in binary outcomes ($n = 5$) or count scores ($n = 6$). Binary outcomes indicated whether participants successfully solved the problem, utilised a specific tool, performed a particular action or whether the solution met predefined criteria. Count scores relate to the number of attempts to solve a problem or the number of different solutions. Other quantitative outcomes are time on task ($n = 3$) and scale scores ($n = 4$).

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Short description and characteristics of the assessment methods.

Assessment method	Studies	Short description	Type	Materials	Data collection	Outcomes/ scores
Air Pressure Tasks	Guevara et al. (2020)[a]	Children are asked to solve a problem with an air pressure apparatus, using provided functional and non-functional objects	Hands-on assignment	Apparatus	Video observation	Binary outcome
Box Tasks	German and Defeyter (2000)[a]	Children are asked to come up with a solution to help a toy bear reach an object, using provided functional and non-functional objects	Hands-on assignment	Craft materials (general), toys	Direct observation	Binary outcome + time
Creating Equal Number Task (CEN)	Leikin (2013); Leikin and Tovli (2014); Tsamir et al. (2010)[a]	Children are asked to make equal number of bottle caps ($N = 10$) on each side of a table in multiple ways.	Hands-on assignment	Craft materials (bottle caps)	Direct observation	Count score
Design Portfolio	Bartholomew et al. (2019)[a]	Children are presented open-ended design challenges, based on popular children's rhymes.	Portfolio assessment	Craft materials (general), photo camera, portfolio	Portfolio	Qualitative coding
Engineering Preschool Children Observation Tool (EPCOT)	Anggoro et al. (2021)[a]	Children are asked to create a solution to a character's problem, based on the concepts of light and shadow.	Hands-on assignment	NA	Video observation	Qualitative coding
Mathematical Problem-Solving Measure	Vessonen et al. (2024)[a]	Children are asked to solve 21 mathematical problems presented in pictures or materials	Oral examination	Pictures, crafts materials, toys	Individual interview	Binary outcome + count score
Metacognitive Skill Tasks	Annevirta and Vauras (2006)[a]	Children are asked to build constructions with crafts materials and Lego bricks, meeting given criteria	Hands-on assignment	Craft materials (general), building bricks (Lego)	Video observation	Qualitative coding + count score

(Continues)

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Assessment method	Studies	Short description	Type	Materials	Data collection	Outcomes/ scores
Open-Ended Problem-Solving Construction Task	Shechter et al. (2021)[a]	Children are asked to build a bridge with Lego bricks over a river (pictorial map), meeting certain conditions	Hands-on assignment	Building bricks (Lego), paper	Direct observation + video observation	Qualitative coding + count score + scale score + binary outcome
Open-Ended Story Problems	Alexander et al. (1994)[a]	Children are asked to finish a story about rescuing a character that has fallen in a deep hole or is locked up in a high tower	Oral examination	Stories	Individual interview	Scale score
Pictorial Multiple Solutions Tasks (PMST 1)	Leikin (2013)[a]; Leikin and Tovli (2014)	Children are asked to provide as many solutions as possible to a visually presented problem	Oral examination	Picture	Individual interview	Count score
Pictorial Multiple Solutions Tasks (PMST 2)	Van Elsen et al. (2024)[a]	Children are asked to provide as many solutions as possible to a visually presented problems	Oral examination	Pictures	Individual interview	Count score
Picture Problem Solving Task	Fusaro and Smith (2018)[a], Lin et al. (2021), Lin et al. (2023)	Children are asked to provide solutions to a visually presented problem and to elaborate on their solution.	Oral examination	Pictures	Individual interview + video observation	Qualitative coding
Problem-Solving Performance Instrument	Nam et al. (2019)[a]	Children are asked to solve a series of mathematical problems, using provided functional and non-functional objects	Hands-on assignment	Craft materials (general)	Direct observation	Binary outcome + Count score
Problem-Solving Scale in Science Education (PSSSE)	Unal and Aral (2014a)[a]; Unal and Aral (2014b); Unal and Saglam (2018)	Children are asked to formulate possible solutions to 16 drawings of science-related problem situation	Oral examination	Pictures	Individual interview	Scale score

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Assessment method	Studies	Short description	Type	Materials	Data collection	Outcomes/ scores
Problem-Solving Skill Scale (PSSS) (Oğuz & Köksal-Akyol, 2015)[a]	Cakir et al. (2021); Demirel and Deretarla Gul (2021)	Children are asked to formulate possible solutions to 18 drawings of problem situation	Oral examination	Pictures	Individual interview	Scale score
Stringent Solution-Standard Tasks	Hong and Milgram (1991)[a]	Children are asked to formulate as many solutions to four problems in real life set-up	Hands-on assignment	Consumer goods (general)	Individual interview	Count score
Tool Improvisation Tasks	Parsonson and Baer (1978)[a]	Children are asked to solve three problems, using provided functional and non-functional objects	Hands-on assignment	Craft materials (general), tools	Direct observation	Count score
Toy-Retrieval Tasks	Gardiner et al. (2012)[a]	Children are asked to get a toy out of a tube, using provided functional and non-functional objects	Hands-on assignment	Apparatus	Direct observation	Binary outcome

[a]Original study in which the assessment method was developed or first described.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Steps of the problem-solving process addressed in the assessments.

Assessment method	Studies	Exploring	Understanding	Representing	Formulating	Planning	Executing	Monitoring	Reflecting	Total (*N*)
Air Pressure Tasks	Guevara et al. (2020)[a]						x			1
Box Tasks	German and Defeyter (2000)[a]						x			1
Creating Equal Number Task (CEN)	Leikin (2013); Leikin and Tovli (2014); Tsamir et al. (2010)[a]				x		x			2
Design Portfolio	Bartholomew et al. (2019)[a]			x	x	x	x	x	x	6
Engineering Preschool Children Observation Tool (EPCOT)	Anggoro et al. (2021)[a]	x	X		x	x	x	x	x	7
Mathematical Problem-Solving Measure	Vessonen et al. (2024)[a]						x			1
Metacognitive Skill Tasks	Annevirta and Vauras (2006)[a]					x	x	x	x	4
Open-Ended Problem-Solving Construction Task	Shechter et al. (2021)[a]	x	x		x	x	x	x	x	7
Open-Ended Story Problems	Alexander et al. (1994)[a]				x					1
Pictorial Multiple Solutions Tasks (PMST 1)	Leikin (2013)[a]; Leikin and Tovli (2014)				x					1
Pictorial Multiple Solutions Tasks (PMST 2)	Van Elsen et al. (2024)[a]				x					1

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 (Continued)

Assessment method	Studies	Exploring	Understanding	Representing	Formulating	Planning	Executing	Monitoring	Reflecting	Total (*N*)
Picture Problem Solving Task	Fusaro and Smith (2018)[a], Lin et al. (2021), Lin et al. (2023)				x					1
Problem-Solving Performance Instrument	Nam et al. (2019)[a]						x			1
Problem-Solving Scale in Science Education (PSSSE)	Unal and Aral (2014a)[a]; Unal and Aral (2014b); Unal and Saglam (2018)				x					1
Problem-Solving Skill Scale (PSSS) (Oğuz & Köksal- Akyol, 2015)[a]	Cakir et al. (2021); Demirel and Deretarla Gul (2021)				x					1
Stringent Solution-Standard Tasks	Hong and Milgram (1991)[a]				x					1
Tool Improvisation Tasks	Parsonson and Baer (1978)[a]				x		x			2
Toy-Retrieval Tasks	Gardiner et al. (2012)[a]					x	x			2
	Total (*N*)	2	2	1	12	5	11	4	4	

[a]Original study in which the assessment method was developed or first described.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 STEM disciplines addressed in the assessments.

Assessment method	Studies	Science	Technology	Engineering	Mathematics	Total (*N*)
Air Pressure Tasks	Guevara et al. (2020)[a]	x	x			2
Box Tasks	German and Defeyter (2000)[a]		x			1
Creating Equal Number Task (CEN)	Leikin (2013); Leikin and Tovli (2014); Tsamir et al. (2010)[a]				x	1
Design Portfolio	Bartholomew et al. (2019)[a]	x	x	x	x	4
Engineering Preschool Children Observation Tool (EPCOT)	Anggoro et al. (2021)[a]	x		x		2
Mathematical Problem-Solving Measure	Vessonen et al. (2024)[a]				x	1
Metacognitive Skill Tasks	Annevirta and Vauras (2006)[a]			x		1
Open-Ended Problem-Solving Construction Task	Shechter et al. (2021)[a]			x		1
Open-Ended Story Problems	Alexander et al. (1994)[a]		x			1
Pictorial Multiple Solutions Tasks (PMST 1)	Leikin (2013)[a]; Leikin and Tovli (2014)		x			1
Pictorial Multiple Solutions Tasks (PMST 2)	Van Elsen et al. (2024)[a]		x			1
Picture Problem Solving Task	Fusaro and Smith (2018)[a], Lin et al. (2021), Lin et al. (2023)	x	x	x	x	4
Problem-Solving Performance Instrument	Nam et al. (2019)[a]				x	1
Problem-Solving Scale in Science Education (PSSSE)	Unal and Aral (2014a)[a]; Unal and Aral (2014b); Unal and Saglam (2018)	x				1
Problem-Solving Skill Scale (PSSS) (Oğuz & Köksal- Akyol, 2015)[a]	Cakir et al. (2021); Demirel and Deretarla Gul (2021)	x	x			2
Stringent Solution-Standard Tasks	Hong and Milgram (1991)[a]		x			1

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 (Continued)

Assessment method	Studies	Science	Technology	Engineering	Mathematics	Total (*N*)
Tool Improvisation Tasks	Parsonson and Baer (1978)[a]		x			1
Toy-Retrieval Tasks	Gardiner et al. (2012)[a]		x			1
	Total (*N*)	6	11	5	5	

[a]Original study in which the assessment method was developed or first described.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Assessment procedures and types of reliability and validity reported in the studies.

Study	Assessment method	Administrator	Location	Duration (min.)	Format	Interrater reliability	Test–retest reliability	Internal consistency (alpha)	Content validity (face validity)	Construct validity (factor analysis)	External validity	Ecological validity
Alexander et al. (1994)	Open-Ended Story Problems	Researcher	NA	NA	Individually	x						
Anggoro et al. (2021)	Engineering Preschool Children Observation Tool (EPCOT)	Researcher	NA	NA	Classical	x						
Annevirta and Vauras (2006)	Metacognitive Skill Tasks	Researcher	NA	35–45	Individually	x		x				
Bartholomew et al. (2019)	Design Portfolio	Researcher	NA	Ca. 60	Individually	x						
Cakir et al. (2021)	Problem-Solving Skill Scale (PSSS)	Researcher	NA	NA	NA		x	x	x	x		
Demirel and Deretarla Gul (2021)	Problem-Solving Skill Scale (PSSS)	Researcher	NA	NA	NA		x	x	x	x		
Fusaro and Smith (2018)	Picture Problem Solving Task	Researcher	NA	NA	Individually	x					x	
Gardiner et al. (2012)	Toy-Retrieval Tasks	Researcher	Quiet room	Unlimited	Individually							
German and Defeyter (2000)	Box Tasks	Researcher	Participant's classroom	Max. 180	Individually	x					x	
Guevara et al. (2020)	Air Pressure Tasks	Researcher	Quiet room	20–25	In pairs							
Hong and Milgram (1991)	Stringent Solution-Standard Tasks	NA	NA	Unlimited	Individually							

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 (Continued)

Study	Assessment method	Administrator	Location	Duration (min.)	Format	Interrater reliability	Test–retest reliability	Internal consistency (alpha)	Content validity (face validity)	Construct validity (factor analysis)	External validity	Ecological validity
Leikin (2013)	Creating Equal Number Task (CEN); Pictorial Multiple Solutions Tasks (PMST 1)	NA	NA	NA	NA							
Leikin and Tovli (2014)	Creating Equal Number Task (CEN); Pictorial Multiple Solutions Tasks (PMST 1)	NA	NA	NA	NA							
Lin et al. (2021)	Picture Problem Solving Task	Graduate students	NA	NA	Individually	x						
Lin et al. (2023)	Picture Problem Solving Task	Graduate students	Quiet Room	10–15	Individually	x						
Nam et al. (2019)	Problem-Solving Performance Instrument	Teacher	NA	Ca. 25	NA	x			x			
Parsonson and Baer (1978)	Tool Improvisation Tasks	Researcher	Quiet room	Ca. 15	Individually	x						
Shechter et al. (2021)	Open-Ended Problem-Solving Construction Task	Researcher	Quiet room	Ca. 20	Individually	x			x			
Tsamir et al. (2010)	Creating Equal Number Task (CEN)	Researcher	Participant's classroom	NA	Individually							

(Continues)

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Eisen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 (Continued)

Study	Assessment method	Administrator	Location	Duration (min.)	Format	Interrater reliability	Test–retest reliability	Internal consistency (alpha)	Content validity (face validity)	Construct validity (factor analysis)	External validity	Ecological validity
Unal and Aral (2014a)	Problem-Solving Scale in Science Education (PSSSE)	NA	Quiet room	10–15	Individually	x	x	x	x	x		
Unal and Aral (2014b)	Problem-Solving Scale in Science Education (PSSSE)	NA	NA	NA	NA		x	x		x		
Unal and Saglam (2018)	Problem-Solving Scale in Science Education (PSSSE)	NA	NA	NA	NA		x	x				
Van Elsen et al. (2024)	Pictorial Multiple Solutions Tasks (PMST 2)	Teacher	Participant's classroom	Ca. 3	Individually	x						x
Vessonen et al. (2024)	Mathematical Problem-Solving Measure	Researcher	Quiet room	Ca. 30	Individually			x				
Information provided: *N* (%)		18 (75)	10 (42)	13 (54)	17 (71)	13 (54)	5 (21)	7 (29)	5 (21)	4 (17)	2 (8)	1 (4)
NOT provided: *N* (%)		6 (25)	14 (58)	11 (46)	7 (29)	11 (46)	19 (79)	17 (71)	19 (79)	20 (83)	22 (92)	23 (96)

Note: This table only presents psychometric properties that were provided by at least one of the included studies.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

An illustrative example of a hands-on problem-solving assignment is the Open-Ended Problem-Solving Construction Task (Shechter et al., 2021). To assess preschoolers' problem-solving competencies, a map of a river was placed on a table and the participants were challenged to build a bridge over the river with standard Lego (Duplo) building bricks, meeting four requirements: (1) the bridge must be structurally stable, (2) the bridge must connect both riverbanks, (3) the bridge must not have any pillar in the water and (4) the bridge must be high enough to allow a paper boat to pass underneath.

The Tool Improvisation Tasks (Parsonson & Baer, 1978) is another example of a hands-on problem-solving assessment method. Participants were challenged to hammer wooden pegs in a peg board. Instead of providing participants with a hammer, the children could choose between seven functional (e.g. a brick) and 10 non-functional items (e.g. a styrofoam hammer) to pound down the pegs. Some of the items could be combined to effective tools (e.g. by putting a rod in a drilled block). In a similar challenge required participants to find effective containers to carry 80 marbles. Again, participants could choose between functional (e.g. a glove) and non-functional items (e.g. a torn bag, recipients that were too small) or could combine items to create useful containers (e.g. by blocking the hole of a bottomless styrofoam cup).

Oral examinations

Most methods that are based on oral examinations, make use of pictures to illustrate problem situations ($n=6$). In the Open-Ended Story Problems (Alexander et al., 1994), the problems were presented as a story. In most cases, the children were asked to come up with multiple solutions to the presented problems. The Mathematical Problem-Solving Measure (Vessonen et al., 2024) required children to solve mathematical problems presented in pictures and materials. Oral examinations typically result in count scores ($n=3$) or scale scores ($n=2$). Some items in the Mathematical Problem-Solving Measure were scored binary (whether or not the children solved the problem) and the Picture Problem Solving Task (Fusaro & Smith, 2018) was processed by qualitative coding.

A typical example of an oral examination is Leikin's Pictorial Multiple Solution Task (PMST; 2013). A black and white picture of an anthropomorphic kitten who can not reach its cap on a high shelf was presented to the children. The picture contains some objects that can be of help, such as a chair, a table, and a stick. Children were asked to come up with as many possible ways to get the cap from the shelf. The answers were scored for fluency (the number of different solutions), flexibility (the number of different strategies employed in the formulated solutions) and originality (a score for unusual or original solutions). This way of scoring solutions is based on the Torrance Test of Creative Thinking (Torrance, 1974) and was also applied in the Pictorial Multiple Solution Task developed by Van Elsen et al. (2024) and the Open-Ended Story Problems (Alexander et al., 1994).

Problem-Solving Skills Scale (PSSS, Cakir et al., 2021; Demirel & Deretarla Gul, 2021) and Problem-Solving Scale in Science Education (PSSSE, Unal & Aral, 2014a) follow a similar approach as the Pictorial Multiple Solution Tasks, but result in scale scores. In the Problem-Solving Skills Scale, participants were asked to formulate multiple solutions for 18 depicted problems. A 5-point scale was used to indicate the number of formulated solutions, ranging from 0 (no solution) to 4 (more than three solutions). In the Problem-Solving Scale in Science Education, children were asked to formulate a solution to 16 depicted problems related to life sciences, health sciences and physical sciences. The answers were scored on a 4-point scale indicating whether (1) the child did not provide a solution or a wrong solution; (2) the answer had a false focus; (3) the answer did not include a next step; or (4) the answer was right and complete.

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Design portfolio

Finally, Design Portfolio (Bartholomew et al., 2019) makes use of a portfolio to assess preschoolers problem-solving competencies. Children were asked to design solutions to problems that appear in popular children's rhymes (Itsy Bitsy Spider, Baa Baa Black Sheep, and Little Boy Blue). A worksheet was developed for each problem on which the children could draw their ideas at the different phases of the problem-solving process. In addition, the children were encouraged to create a prototype of their solution using crafts materials. At the end, portfolios and prototypes were photographed and all worksheets and prototypes were coded holistically.

Scope of the assessment methods

To address the second research question, we first analyse the problem-solving steps included in the identified methods. Next, we examine which STEM disciplines are emphasised within these methods. Table 3 provides an overview of the problem-solving steps covered by the assessment methods. Table 4 shows the STEM disciplines addressed.

Steps of the problem-solving process

Nine identified methods focus on one step of the problem-solving process. Seven of these methods solely address Formulating possible solutions. This is the case for all oral examinations, except for the Problem-Solving Performance Instrument (Nam et al., 2019) which assesses Executing. Three methods capture nearly the entire problem-solving process: Design Portfolio (Bartholomew et al., 2019), Mathematical Problem-Solving Measure (Vessonen et al., 2024) and Open-Ended Problem-Solving Construction Task (Shechter et al., 2021). These three methods all contain hands-on problems and make use of tangible objects. Overall, Formulating ($n=12$) and Executing ($n=11$) are the most assessed steps of the problem-solving process in the methods included in this review. Exploring ($n=2$), Understanding ($n=2$) and Representing ($n=1$) are assessed least.

STEM disciplines

Regarding the STEM disciplines, most methods focus exclusively on one STEM discipline ($n=13$). Five studies are concerned with more than one STEM discipline of which two address all STEM disciplines: Design Portfolio's (Bartholomew et al., 2019) and Picture Problem-Solving Task (Fusaro & Smith, 2018).

Technology is the most addressed discipline ($n=11$). Assessment methods related to technology typically require children to retrieve hard-to-reach objects, as in the Box task (German & Defeyter, 2000), Stringent Solution-Standard Tasks (Hong & Milgram, 1991), Toy Retrieval Task (Gardiner et al., 2012). Other methods require children to generate tool-based solutions for object retrieval, such as the Open-Ended Story Problems (Alexander et al., 1994) and Pictorial Multiple Solution Tasks (Leikin, 2013; Van Elsen et al., 2024). The Picture Problem-Solving Task (Fusaro & Smith, 2018) requires children to come up with ideas to solve a problem such as getting a strawberry out of an ice cube or finding a way to discriminate between broccoli seeds and carrot seeds that are dropped on the ground and got mixed up. In the Tool Improvisation Tasks (Parsonson & Baer, 1978), children are challenged to use unconventional materials to replace conventional tools.

Three out of four methods that include engineering are hands-on and require children to build or construct solutions to a problem. For the Engineering Preschool Children Observation Tool (EPCOT, Anggoro et al., 2021), children were asked to create a solution for Panda who wants to play outside when it is too hot and bright. In the Metacognitive Skill Task (Annevirta & Vauras, 2006), children are asked to build a fence to protect inhabitants of an island against a dangerous wolf and, in another task, to separate the island in three parts using building bricks (Lego) and crafts materials. As discussed earlier, the Open-Ended Problem-Solving Construction Task (Shechter et al., 2021) required children to build a bridge over a river using standard Lego (Duplo) building bricks. The Picture Problem Solving Task (Fusaro & Smith, 2018) is the only method including engineering in which data is collected by oral examination.

Science-oriented problem-solving is most often combined with technology or engineering. In the Air Pressure Task (Guevara et al., 2020), children had to build an air pressure mechanism with syringes and tubes of different sizes and shapes to reach a predefined goal. Solving the problem requires applying scientific and technological knowledge and skills. The Air Pressure Task is the only included science-related method, which is hands-on. The other three science-related methods are based on oral examination.

Among the large number of well-structured mathematical problems (e.g. mathematical operations, mathematical word problems, seriation problems, unambiguous classification problems) that came up during screening, we found three methods that made use of ill-structured mathematical problems or problems with more than one correct answer: Creating Equal Number Task (CEN, Tsamir et al., 2010), Mathematical Problem-Solving Measure (Vessonen et al., 2024), and Problem-Solving Performance Instrument (Nam et al., 2019). The Creating Equal Number Task (Tsamir et al., 2010) required children to distribute a number of bottle caps so that there were an equal number of bottle caps on each side of a table. After the first rearrangement was done, the children were asked whether they could think of another way to rearrange the bottle caps. Not only the solutions the children came up with were assessed, but also the strategies they applied to rearrange the bottle caps. In addition to well-structured mathematical problems related to categorising, comparing, patterning, and numbering, the Problem-Solving Performance Instrument (Nam et al., 2019) requires children to compare the length of lines using craft materials (e.g. pieces of paper, paper clips, strings, pipe cleaners). After finding a solution, children are asked to think of multiple ways to find out which line is longer. Likewise, the Mathematical Problem-Solving Measure (Vessonen et al., 2024) includes both open-ended problems in addition to closed ones. For the open-ended problems, children were asked to sort different objects into groups in different ways and to explain the criteria they used for sorting. In another item, they were asked to organise different coloured ice cream balls in a paper cone in multiple ways or to match shoes with monsters in multiple combinations.

The Design Portfolio (Bartholomew et al., 2019) explicitly address problem solving in integrated STEM. As mentioned above, children are asked to design a solution to problem that emerges in popular children's rhymes. This requires them to think outside the individual STEM-boxes and find their way in a barely limited problem space.

Assessment procedures and psychometric properties of the assessment methods

The third research question relates to the procedures and psychometric properties of the test. This section presents the administration procedures (including format, administrator, location and duration) and reported types and metrics related to reliability and validity. Table 5

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

provides an overview of key aspects concerning procedures and psychometric properties of the assessment methods as reported in the studies.

Assessment procedures

In most studies, the assessment was administered by an external assessor, not affiliated with the school of the participants, including researchers, ($n=13$), graduate students ($n=2$) and research assistants ($n=1$). In two studies, the preschoolers' own teacher conducted the assessment (Nam et al., 2019; Van Elsen et al., 2024). Six studies did not specify who administered the assessment.

In seven of the 10 studies that reported the location, the assessment took place in a quiet, separate classroom. The three other assessments were conducted in the participant's own classroom. More than half of the studies (58%) did not provide any information on the assessment location.

The reported duration of assessment administration ranged from 3 to 180 min. Approximately half of the studies that provided information on duration reported a time frame between 10 and 30 min. Three studies indicated a duration of more than 30 min (Annevirta & Vauras, 2006; Bartholomew et al., 2019; German & Defeyter, 2000). In two studies, participants were not subject to any time restrictions (Gardiner et al., 2012; Hong & Milgram, 1991).

In all but two studies, the assessment was administered individually. Guevara et al. (2020) assessed the children in pairs and in the study of Anggoro et al. (2021) the assessments were carried out classically.

Psychometric properties of the assessment methods

Six included studies did not report any issues or statistics concerning reliability and validity. In more than half of the studies (54%), interrater reliability was assessed and reported. In these studies, interrater reliability was analysed by comparing scores or codes provided by multiple raters and calculating Cohen's kappa ($n=4$), percentage of agreement ($n=3$) or intraclass correlation (ICC, $n=2$). Unal and Aral (2014a) reported Wilcoxon Signed Rank Test Scores as a measure for interrater reliability and German and Defeyter (2000) stated that there were no disagreements between two raters. The reported kappa values range from 0.60 (Bartholomew et al., 2019) to 0.81 (Lin et al., 2021), corresponding respectively to moderate and strong agreement (McHugh, 2012). The percentage of agreement ranged from 67 to 100% (Parsonson & Baer, 1978). The reported ICC values range from 0.69 to 1 (Van Elsen et al., 2024). Van Elsen et al. (2024) observed better interrater agreement between four teachers (ICC = 1) than between four researchers (ICC = 0.69–0.89). ICC values above 0.60 are generally considered fair (Cicchetti, 1994), while values exceeding above 0.75 are regarded as desirable (Koo & Li, 2016). It should be noted, however, that different forms of ICC exist. Test–retest reliability was discussed in five studies. The reported correlations between tests ranged from 0.60 (Cakir et al., 2021) to 0.96 (Unal & Aral, 2014a).

Internal consistency was reported in seven studies. Cronbach's alpha ranged from 0.70 to 0.90 (Shechter et al., 2021). Cronbach's alpha around or above 0.70 is generally considered desirable (Taber, 2018). Vessonen et al. (2024) reported that internal consistency of the mathematical problem-solving measure was acceptable for children aged 4 and 5, but fell below acceptable levels for 3-year-olds. Therefore, the youngest age group was not included in the statistical analyses.

Content validity was established by involving two (Nam et al., 2019), six (Shechter et al., 2021), or ten (Unal & Aral, 2014a) experts in preschool education and research to

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

evaluate the instrument (face validity). Two studies calculated the Content Validity Index which ranged from 0.96 to 0.99, although details on how these figures were obtained and how they should be interpreted are lacking (Cakir et al., 2021; Demirel & Deretarla Gul, 2021). The remaining studies ($n=19$) did not report on content validity. Four studies examined and reported the factor structure of a scale (construct validity) with exploratory factor analysis (EFA). However, only (Unal & Aral, 2014a) and (Unal & Aral, 2014b) provided an in-depth discussion of the results.

External validity was addressed by German and Defeyter (2000), who discussed results from two replication studies, and Fusaro and Smith (2018), who briefly mentioned the generalisability of their findings in the limitations section. Only Van Elsen et al. (2024) explicitly addressed ecological validity by noting that teachers were involved in the assessment process.

DISCUSSION

Problem-solving is a key competence for lifelong learning (European Commission, 2019). Learning to solve problems takes time (Molnár et al., 2013). Therefore, teaching children how to solve problems best starts from preschool onwards (Burns et al., 2025; Lile Diamond, 2018; Rogers, 2004). Problem-solving is a fundamental component of STEM education and STEM activities offer numerous opportunities for young children to engage with ill-structured problems and develop essential problem-solving competencies (Priemer et al., 2020). To monitor preschoolers' development of problem-solving competencies and evaluate the effectiveness of classroom interventions, researchers and educators need reliable and valid assessment methods. Up to now, an overview of methods that are developed, studied or used in scientific research to assess problem-solving competencies within the context of early STEM education research was lacking. With this scoping review, we aimed at providing such an overview and analysed the characteristics (RQ 1), scope (RQ 2), procedures and psychometric properties (RQ 3) of these methods. The results can be used by researchers and educators to select, implement or develop reliable and valid methods for assessing STEM-based problem solving.

A systematic search in Web of Science, Scopus and Eric resulted in 24 eligible studies in which 18 different methods related to problem-solving were identified. We found that the typical method is based on direct observation of hands-on problem assignments or is based on oral examination, resulting in binary outcomes or count scores and focusses on a single step of the problem-solving process in a single STEM discipline. In addition, we found that the assessments are generally conducted individually, in a separate classroom, by an external researcher, and last about 10 to 30 min. Finally, it became evident that most assessment instruments had been developed on an ad hoc basis, with limited consideration given to reliability and validity. In the following sections, we elaborate on the main findings and discuss some limitations of our study. We end with implications and recommendations for research and practice.

Characteristics of assessment methods

Our scoping review of the scientific literature revealed that researchers most commonly use either observations of hands-on activities or oral examinations to assess the problem-solving competencies of preschoolers. Both methods have their merits. Observation allows teachers and researchers to record what preschoolers effectively do in a real-world context. It does not depend heavily on a certain level of verbal skills, it provides opportunities to record both verbal and non-verbal behaviour, and it offers opportunities to map social interactions during performance (Azevedo, 2009). Brassard and Boehm (2007) argue that observation of meaningful, unstructured tasks is an essential part of any assessment of young children.

Oral examination, on the other hand, can be used to assess (meta-)cognitive processes and competencies that are difficult to observe, such as coming up with multiple solutions or reflecting. Engaging in dialogue with children can elicit reflective thinking (Stoll et al., 2012) and allows researchers and educators to verify whether preschoolers have correctly understood the problem and supports the interpretation of their observed behaviour (Snow & Van Hemel, 2008).

Hands-on assignments often make use of easily available and familiar crafts materials (e.g. cardboard, tape, glue, bottle caps, pipe cleaners) or popular toys (e.g. Lego building bricks). This allows children to build on their previous experience with the material's properties and possible uses. When unfamiliar materials are being used, it is suggested to allow them more or even unlimited time to explore the materials and their features (Gardiner et al., 2012). Another benefit of using ubiquitous materials for the assessment is that this makes the assessment method applicable in all preschool classrooms, which facilitates the use of the methods in multiple studies and daily practice.

The use of visual representations, such as pictures depicting problem situations, emerged as a common practice in problem-solving assessments. Given their developing language skills, pictures can help preschoolers correctly interpret the problem. As stated by Polya (1945), a problem that is not well understood, is impossible to solve. Whereas older students can be encouraged to create their own representation of the problem (OECD, 2013), preschoolers may benefit from visual support.

Most assessment methods, whether observation-based or based on oral examination, result in quantitative outcomes (binary score, count score, or scale score). Although five studies applied qualitative coding, the resulting categories were in most cases used for quantitative analysis. Among the reviewed studies, only the Design Portfolios (Bartholomew et al., 2019) included qualitative outcomes such as children's drawings and photographs. In addition to other methods, portfolio assessment can be a powerful tool for researchers and educators who want to monitor the development and learning of young children (Seitz & Bartholomew, 2008).

Scope of the assessment methods

Altogether, the assessment methods included in this review covered all the steps of the PISA problem-solving framework (OECD, 2013). However, our review reveals that the individual methods generally focus on one or at most two steps of the problem-solving process. We only found three methods that include roughly the entire problem-solving process. However, problem-solving is more than the sum of isolated knowledge and skills (Jonassen, 1997; Maker, 2001; Martín-Páez et al., 2019). It requires metacognitive and practical skills as well as certain attitudes and the willingness to engage with the problem that cannot be assessed with simple tasks (Mayer, 1998; OECD, 2013). To become proficient problem-solvers, children must learn to integrate the different steps of the problem-solving process (Polya, 1945). This highlights the need for a more integrated approach to assessing problem-solving competencies in preschool education.

The same applies for the assessed STEM disciplines. Although STEM is often used as an umbrella term for educational activities within the four STEM disciplines (Martín-Páez et al., 2019), in the context of solving ill-structured STEM problems, STEM is best conceptualised as an integrative approach to assessing and teaching science, technology, engineering and mathematics (Wan et al., 2021). Ill-structured problems are embedded in their context and require knowledge and processes from more than one STEM discipline. However, most

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

assessment methods included in this review focus on one discipline. We found five methods that address more than one STEM discipline, but only the Design Portfolio (Bartholomew et al., 2019) is explicitly embedded in an integrated STEM approach. These results are in line with the concern of Douglas et al. (2020) that educational assessment in STEM is often directed at domain-specific knowledge and skills. In formal education, segregated STEM problems (i.e. problems involving only knowledge and skills from one STEM discipline) are typically well-structured, resulting in one best and known solution and are associated with direct instruction and a focus on content (Nadelson & Seifert, 2017). According to Martín-Páez et al. (2019), an isolated approach to the STEM disciplines is far removed from authentic, ill-structured STEM problems. Therefore, there remains a need for assessment methods that incorporate ill-structured problems and integrate the STEM disciplines.

Although initially underrepresented (Kelley & Knowles, 2016), technology and engineering have gained attention in recent years and are now seen as an important component of science and mathematics (Bryan & Guzey, 2020). Among the assessment methods included in this review, 46% address technology either exclusively or in combination with other STEM domains, which is substantially higher than the 21–25% observed for the other STEM domains. Of all STEM disciplines, technology is perhaps the most visible aspect of real-life problems (Aydin Gunbatar et al., 2022). Technology is also the most 'tangible' STEM discipline, making it more accessible to preschoolers. Technology is instrumental for assessing and teaching other STEM disciplines.

Engineering, on the other hand, is somewhat less represented in the identified assessment methods in this review (21%). According to Martín-Páez et al. (2019), engineering can bring all STEM-disciplines together. Engineering is strongly related to design. Two studies included in this review explicitly relate problem solving to the engineering design process (Anggoro et al., 2021) or design thinking (Bartholomew et al., 2019). Both methods cover nearly the entire problem-solving process. Although design is not part of the STEM-acronym, engineering design is 'the interdisciplinary glue' between all STEM disciplines (Tank et al., 2018, p. 183) and has the potential to integrate all steps of the problem-solving process (English, 2023).

Assessment procedures and psychometric properties of assessment methods

Most assessments described in the included studies were conducted individually in a quiet, controlled environment under the supervision of a researcher or assistant. This aligns with the observation of Rhodes et al. (2024) that problem-solving competencies in preschoolers are typically assessed in controlled settings, removed from the authentic classroom context. However, Brassard and Boehm (2007) and DeLuca et al. (2020) emphasise the importance of embedding preschool assessment in authentic learning environments.

Consistent with the findings of Cutumisu et al. (2019), most studies implemented a new assessment method tailored to their specific needs, while giving limited consideration to key issues related to the method's reliability or validity. In line with Barghaus et al. (2022) and Burns et al. (2025), our review reveals that limited attention is given to the psychometric properties of instruments that are used in research to assess problem-solving competencies in preschoolers. A quarter of the reviewed studies lack any reporting on the reliability and validity of the assessment instruments used. Although half of the studies report statistics on interrater reliability, other aspects related to reliability and validity are often overlooked.

The reported psychometric properties generally fall within the acceptable to strong range, with Cronbach's alpha values around or above 0.70 (Taber, 2018), ICC values around or above 0.75 (Koo & Li, 2016) and Cohen's Kappa values around or above 0.80

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(McHugh, 2012), although some results approach the lower threshold. In studies where inadequate internal consistency or reliability was identified, the authors either excluded part of the data (Vessonen et al., 2024) or refined the instrument further (Anggoro et al., 2021).

Consistent with Burns et al. (2025), we found that reliability and validity were only systematically examined for the Problem Solving Scale in Science Education (PSSSE) developed by Unal and Aral (2014a). Here we add that Cakir et al. (2021) and Demirel and Deretarla Gul (2021) also address multiple aspects concerning validity and reliability in their respective studies, albeit in a more limited and less systematic manner. These findings underscore the need for reliable and valid problem-solving assessment methods in early childhood, as recently emphasised by Burns et al. (2025).

Limitations

For this review, we followed the rigorous JBI and PRISMA standards for scoping reviews. Nevertheless, this scoping review is not without limitations. We only included peer reviewed journal articles. Although scoping reviews allow for the inclusion of grey literature (e.g. Pollock et al., 2021), it was not possible to provide a systematic and exhaustive overview of problem-solving assessments based on grey literature due to limited accessibility and variety of this type of evidence. In addition, grey literature does not always meet the required quality criteria concerning reliability and validity. As a result, this review may be subject to publication bias since we may have missed interesting assessment methods that were not published in scientific journals or indexed by Scopus, Web of Science, and Eric.

There is also a potential risk of bias due to the study selection. For this review, we only included studies that used assessment methods involving ill-structured or open-ended problems. However, the classification of problems in ill-structured and well-structured problems is not unambiguous. As Jonassen (2010) points out, all problems can be situated on a continuum of structuredness. Consequently, we had to make informed judgements about the inclusion of studies that focused on problems situated midway along the continuum. To ensure consistency, we compiled a list of types of problems encountered during the initial screening round. This list was then discussed with the research team, and decisions about inclusion and exclusion were made by consensus. The resulting list was subsequently used during the second round of screening.

Finally, it is worth noting that we deviated from our initial protocol in two ways. In scoping reviews, study selection and data extracting is often an iterative process (Aromataris & Munn, 2020; Levac et al., 2010; Pollock et al., 2021). Given the diversity of problems used to assess problem-solving competencies, we slightly adjusted the eligibility criteria by specifying eligible problem types and conducted a second round of screening. After the first round of screening, we also decided to adjust the stopping criteria during title and abstract screening. Rather than stopping after 50 consecutive rejections as initially planned and recommended by Ros et al. (2017), we chose to continue until 100 consecutive rejections to minimie the risk of overlooking relevant studies.

Implications for practice and research

The current review illustrates the tendency to approach problem-solving assessment in a segregated STEM structure. In most K-12 education systems, STEM is discipline-based, while STEM problems are not (Nadelson & Seifert, 2017). Rethinking the historically rooted division of STEM disciplines can be conducive to assessing and teaching competencies that are needed for solving authentic, ill-structured STEM-problems. According to Nadelson

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and Seifert (2017), researchers and educators also need the knowledge and willingness to introduce ill-structured or open-ended problems in the classroom and feel comfortable with the associated uncertainties. Allowing children to engage in ill-structured problems means giving up discipline-boundaries and control of the problem space.

We also note that there is still a need for assessment methods that can be administered multiple times and can be used for monitoring the development of preschoolers' problem-solving competencies in longitudinal studies, longitudinal studies, or studies with a pretest, post-test, retention test design. Most methods included in this review can only be administered ones. Problems previously encountered and solved, are no longer real problems because preschoolers can build on prior experience and are expected to already know a solution or a solution strategy.

The study revealed a notable lack of assessment instruments for which reliability and validity have been thoroughly examined. This impedes research on the development of children's problem-solving competencies and the effectiveness of problem-solving interventions. To move the field forward, there is a clear need for developmentally appropriate, reliable, valid and authentic assessment methods that reflect children's real-world problem-solving competencies. We also found that researchers often develop ad hoc assessment tools tailored to the specific needs of their studies. Reusing existing instruments can facilitate the validation of these instruments. Alternatively, new instruments can be developed in accordance with the guidelines outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

Among the 18 methods we found, Bartholomew et al.'s (2019) Design Portfolio stands out in several ways. First, it captures nearly the entire problem-solving process by asking children to draw their ideas (Representing and Formulating), plan their actions (Planning), effectively build the design (Executing), and evaluate if the design worked (Monitoring and Reflecting). Although Exploration and Understanding are not included in the assessment, they can easily be added since the assignments already elicits exploratory behaviour by asking children explicitly to explore the materials. Preschoolers' Understanding can be derived from their drawings and follow-up questions. Second, the Design Portfolio explicitly addresses STEM in an integrated way. Third, it makes use of tangible materials and follow-up questions to get a better understanding of what has been observed. Although the Design Portfolio as such may not be directly applicable as an assessment method in preschool practice due to the necessity of qualitative coding, it offers a promising mix of elements that can be further refined and developed into effective assessment tools.

Observing children solve engineering design problems and engaging in dialogue with them seems another promising way to assess their problem-solving competencies. Engineering design problems are motivating, open-ended, hands-on problems that require children to construct or craft solutions by going through all steps of the problem-solving process (Cunningham et al., 2018; Dubosarsky et al., 2018; Isabelle et al., 2021), incorporating all STEM disciplines (Tank et al., 2018). These types of problems closely align with preschoolers' play and daily activities in preschool classrooms, which makes them suitable for authentic educational assessment.

CONCLUSION

This scoping review provides an overview of the characteristics, scope, procedures and psychometric properties of assessment methods that are developed, studied or used in scientific research to assess preschoolers' problem-solving competencies. The findings support researchers and educators in selecting, implementing or developing reliable and valid instruments, ultimately contributing to a more comprehensive understanding of these competencies. Our review of the scientific literature revealed that there is still a need for scientifically substantiated, reliable, and valid methods to assess STEM-based problem solving in an integrated way. The review also highlights the need for increased attention to and transparency concerning the procedures and psychometric properties of assessment methods. Based on our findings, we propose combining observation of hands-on problem assignments with oral examination as an age-appropriate method for assessing preschoolers' problem-solving competencies.

AUTHOR CONTRIBUTIONS

Joke Torbeyns: Conceptualization; validation; writing – review and editing; supervision; methodology; funding acquisition. **Sven De Maeyer:** Conceptualization; validation; supervision; writing – review and editing; methodology; funding acquisition. **Joris Van Elsen:** Conceptualization; investigation; funding acquisition; writing – original draft; methodology; validation; visualization; formal analysis; project administration; data curation.

ACKNOWLEDGEMENTS

The authors would like to thank the librarians of the University of Antwerp for their help in selecting databases and refining the search string.

FUNDING INFORMATION

This study was supported by the Research Foundation—Flanders (Fonds Wetenschappelijk Onderzoek—Vlaanderen) [grant number 1S03525N].

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare that are relevant to the content of this article.

DATA AVAILABILITY STATEMENT

All data generated or analysed during this study are included in this published article.

ETHICS STATEMENT

Not applicable.

REGISTRATION

A protocol was pre-registered on Open Science Framework on 1 July 2022 (https://doi.org/10.17605/OSF.IO/QPJAT).

AI USE DECLARATION STATEMENT

During the preparation of this work, the authors used ASReview for screening title and abstract and ChatGPT, Copilot, Claude, DeepL and DeepL Write for writing assistance and translation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. The authors have verified all AI-generated content for accuracy, confirmed that AI is not listed as an author, reviewed language suggestions, verified translations and reviewed the work for bias and ethical considerations.

ORCID

Joris Van Elsen https://orcid.org/0000-0002-9674-2927
Sven De Maeyer https://orcid.org/0000-0003-2888-1631
Joke Torbeyns https://orcid.org/0000-0003-3244-0265

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Endnote

[i] This search string conforms to the syntax of Scopus. The W/5 in the first argument is a proximity operator to allow a distance of five words between both terms.

REFERENCES

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

*Alexander, P. A., Jetton, T. L., White, S. H., Parsons, J. L., Cotropia, K. K., Liu, H.-C., & Ackerman, C. M. (1994). Young children's creative solutions to realistic and fanciful story problems. *Journal of Creative Behavior*, *28*(2), 89–106. https://doi.org/10.1002/j.2162-6057.1994.tb00723.x

*Anggoro, F. K., Dubosarsky, M., & Kabourek, S. (2021). Developing an observation tool to measure preschool children's problem-solving skills. *Education in Science*, *11*(12), 779. https://doi.org/10.3390/educsci11120779

*Annevirta, T., & Vauras, M. (2006). Developmental changes of metacognitive skill in elementary school children. *Journal of Experimental Education*, *74*(3), 195–226. https://doi.org/10.3200/JEXE.74.3.195-226

Aromataris, E., & Munn, Z. (Eds.). (2020). *JBI manual for evidence synthesis*. JBI. https://doi.org/10.46658/JBIMES-20-01

Aydin Gunbatar, S., Ekiz Kiran, B., Boz, Y., & Roehrig, G. H. (2022). A closer examination of the STEM characteristics of the STEM activities published in NSTA journals. *Research in Science & Technological Education*, *42*(4), 952–977. https://doi.org/10.1080/02635143.2022.2121692

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7

Barghaus, K. M., Fantuzzo, J. W., Buek, K., & Gullo, D. F. (2022). Neglected validities: A diagnostic look at the state of early childhood assessment. *Early Childhood Research Quarterly*, *58*, 287–299. https://doi.org/10.1016/j.ecresq.2021.09.007

*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

Bolden, B., DeLuca, C., Kukkonen, T., Roy, S., & Wearing, J. (2020). Assessment of creativity in K-12 education: A scoping review. *Review of Education*, *8*(2), 343–376. https://doi.org/10.1002/rev3.3188

Bransford, J. D., & Barry, S. S. (1993). *The ideal problem solver. A guide for improving thinking, learning, and creativity* (2nd ed.). W. H. Freeman.

Brassard, M. R., & Boehm, A. E. (2007). *Preschool assessment. Principles and practice*. Guilford.

Brownell, W. A. (1942). Problem solving. *Teachers College Record*, *43*(10), 415–443. https://doi.org/10.1177/016146814204301012

Bryan, L., & Guzey, S. S. (2020). K-12 STEM education: An overview of perspectives and considerations. *Education*, *1*(1), 5–15. https://doi.org/10.51724/hjstemed.v1i1.5

Burns, S., Saleem, S., McMullen, E., Falenchuk, O., White, L., Dhuey, E., & Perlman, M. (2025). A systematic review and meta-analysis of approaches to teaching problem-solving skills in early childhood education and care settings: A focus on science, technology, engineering and mathematics activities. *Review of Education*, *13*(2), e70079. https://doi.org/10.1002/rev3.70079

*Cakir, R., Korkmaz, O., Idil, O., & Erdogmus, F. U. (2021). The effect of robotic coding education on preschoolers' problem solving and creative thinking skills. *Thinking Skills and Creativity*, *40*, 100812. https://doi.org/10.1016/j.tsc.2021.100812

Christou, E., Parmaxi, A., & Zaphiris, P. (2025). A systematic exploration of scoping and mapping literature reviews. *Universal Access in the Information Society*, *24*(1), 941–951. https://doi.org/10.1007/s10209-024-01120-3

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clements, D. H., & Sarama, J. (2016). Math, science, and technology in the early grades. *The Future of Children*, *26*(2), 75–94.

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). Springer. https://doi.org/10.1007/978-981-10-8621-2_8

Cutumisu, M., Adams, C., & Lu, C. (2019). A scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology*, *28*(6), 651–676. https://doi.org/10.1007/s10956-019-09799-3

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

*Demirel, B., & Deretarla Gul, E. (2021). The effect of riddles on problem solving skills of 57-66 months old children [article]. *Cukurova University Faculty of Education journal*, *50*(2), 721–748. https://doi.org/10.14812/cufej.943800

Dewey, J. (1910). *How we think*. D. C. Heath & CO. https://www.gutenberg.org/files/37423/37423-h/37423-h.htm

Douglas, K. A., Gane, B. D., Neumann, K., & Pellegrino, J. W. (2020). Contemporary methods of assessing integrated STEM competencies. In C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & L. D. English (Eds.), *Handbook of research on STEM education* (pp. 234–254). Routledge.

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). Springer. https://doi.org/10.1007/978-981-10-8621-2_12

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5), i–113. https://doi.org/10.1037/h0093599

English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education*, *3*(1), 3. https://doi.org/10.1186/s40594-016-0036-1

English, L. D. (2023). Ways of thinking in STEM-based problem solving. *ZDM Mathematics Education*, *55*, 1219–1230. https://doi.org/10.1007/s11858-023-01474-7

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*, 105091. https://doi.org/10.1016/j.jecp.2021.105091

Frensch, P., & Funke, J. (1995). *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates. https://www.researchgate.net/publication/200134353_Complex_Problem_Solving-The_European_Perspective

Funke, J., & Frensch, P. (2007). Complex problem solving: The European perspective – ten years after. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25–47). Lawrence Erlbaum. https://doi.org/10.4324/9781315091938-2

*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

*Gardiner, A. K., Bjorklund, D. F., Greif, M. L., & Gray, S. K. (2012). Choosing and using tools: Prior experience and task difficulty influence preschoolers' tool-use strategies. *Cognitive Development*, *27*(3), 240–254. https://doi.org/10.1016/j.cogdev.2012.05.001

*German, T. P., & Defeyter, M. A. (2000). Immunity to functional fixedness in young children. *Psychonomic Bulletin and Review*, *7*(4), 707–712. https://doi.org/10.3758/BF03213010

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, *13*, 74–83. https://doi.org/10.1016/j.edurev.2014.10.002

*Guevara, M., Rojas Ospina, T., & van Geert, P. (2020). Understanding transfer from a dynamic system approach: Two studies of children using problem-solving tasks. *Cogent Education*, *7*, 1814512. https://doi.org/10.1080/2331186X.2020.1814512

Harlen, W. (2007). Meanings, processes and properties of assessment. In *Assessment of learning* (pp. 11–24). SAGE. https://doi.org/10.4135/9781446214695.n2

*Hong, E., & Milgram, R. M. (1991). Original thinking in preschool children: A validation of ideational fluency measures. *Creativity Research Journal*, *4*(3), 253–260. https://doi.org/10.1080/10400419109534397

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

ITEEA. (2020). Standards for technological and engineering literacy. *The role of technology and engineering in stem education*. https://www.iteea.org/downloadpurchase-stel

Jonassen, D. H. (1997). Instructional design models for well-structured and III-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65–94. https://doi.org/10.1007/BF02299613

Jonassen, D. H. (2010). *Learning to solve problems. A handbook for designing problem-solving learning environments*. Routledge. https://doi.org/10.4324/9780203847527

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

*Leikin, M. (2013). The effect of bilingualism on creativity: Developmental and educational perspectives. *International Journal of Bilingualism*, *17*(4), 431–447. https://doi.org/10.1177/1367006912438300

*Leikin, M., & Tovli, E. (2014). Bilingualism and creativity in early childhood. *Creativity Research Journal*, *26*(4), 411–417. https://doi.org/10.1080/10400419.2014.961779

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

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*(1), 69. https://doi.org/10.1186/1748-5908-5-69

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

*Lin, X., Yang, W., Wu, L., Zhu, L., Wu, D., & Li, H. (2021). Using an inquiry-based science and engineering program to promote science knowledge, problem-solving skills and approaches to learning in preschool children. *Early Education and Development*, *32*(5), 695–713. https://doi.org/10.1080/10409289.2020.1795333

*Lin, X. Y., Yang, W. P., Xie, W. L., & Li, H. (2023). The integrative role of parenting styles and parental involvement in young children's science problem-solving skills. *Frontiers in Psychology*, *14*, 1096846. https://doi.org/10.3389/fpsyg.2023.1096846

Lippard, C. N., Riley, K. L., & Lamm, M. H. (2018). Encouraging the development of engineering habits of mind in prekindergarten learners. In L. English & T. Moore (Eds.), *Early engineering learning* (pp. 19–36). Springer. https://doi.org/10.1007/978-981-10-8621-2_3

Lopes, C. E., Grando, R. C., & D'Ambrosio, B. S. (2017). Experiences situating mathematical problem solving at the core of early childhood classrooms. *Early Childhood Education Journal*, *45*(2), 251–259. https://doi.org/10.1007/s10643-016-0775-0

Lu, C., Macdonald, R., Odell, B., Kokhan, V., Demmans Epp, C., & Cutumisu, M. (2022). A scoping review of computational thinking assessments in higher education. *Journal of Computing in Higher Education*, *34*(2), 416–461. https://doi.org/10.1007/s12528-021-09305-y

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 in Science*, *12*(5), 352. https://doi.org/10.3390/educsci12050352

Maker, C. J. (2001). DISCOVER: Assessing and developing problem solving. *Gifted Education International*, *15*(3), 232–251. https://doi.org/10.1177/026142940101500303

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

Martín-Páez, T., Aguilera, D., Perales-Palacios, F. J., & Vílchez-González, J. M. (2019). What are we talking about when we talk about STEM education? A review of literature. *Science Education*, *103*(4), 799–822. https://doi.org/10.1002/sce.21522

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*(1), 49–63. https://doi.org/10.1023/A:1003088013286

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)*, *22*(3), 276–282.

McMillan, J. (2013). *SAGE handbook of research on classroom assessment*. SAGE Publications, Inc. https://doi.org/10.4135/9781452218649

Mehsen, V., Morag, L., Chesta, S., Cleaton, K., & Burgos, H. (2022). Hot executive function assessment instruments in preschool children: A systematic review. *International Journal of Environmental Research and Public Health*, *19*(1), 95.

Milburn, T. F., Lonigan, C. J., DeFlorio, L., & Klein, A. (2019). Dimensionality of preschoolers' informal mathematical abilities. *Early Childhood Research Quarterly*, *47*, 487–495. https://doi.org/10.1016/j.ecresq.2018.07.006

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, *9*, 35–45. https://doi.org/10.1016/j.tsc.2013.03.002

Moore, T. J., Johnston, A. C., & Glancy, A. W. (2020). STEM integration. A synthesis of conceptual frameworks and dafinitions. 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.

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Nadelson, L. S., & Seifert, A. L. (2017). Integrated STEM defined: Contexts, challenges, and the future. *The Journal of Educational Research*, *110*(3), 221–223. https://doi.org/10.1080/00220671.2017.1289775

*Nam, K. W., Kim, H. J., & Lee, S. (2019). Connecting plans to action: The effects of a card-coded robotics curriculum and activities on Korean kindergartners. *The Asia-Pacific Education Researcher*, *28*(5), 387–397. https://doi.org/10.1007/s40299-019-00438-4

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

OECD. (2004). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. OECD.

OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD. https://doi.org/10.1787/9789264190511-en

Oğuz, V., & Köksal- Akyol, A. (2015). Problem Çözme Becerisi Ölçeği (PÇBÖ) geçerlik ve güvenirlik çalışması [Problem-Solving Skills Scale (PSSS) validity and reliability]. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, *44*(1), 105–122.

O'Reilly, C., Devitt, A., & Hayes, N. (2022). Critical thinking in the preschool classroom – a systematic literature review. *Thinking Skills and Creativity*, *46*, 101110. https://doi.org/10.1016/j.tsc.2022.101110

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

*Parsonson, B. S., & Baer, D. M. (1978). Training generalized improvisation of tools by preschool children. *Journal of Applied Behavior Analysis*, *11*(3), 363–380. https://doi.org/10.1901/jaba.1978.11-363

Philpott-Robinson, K., Johnson, T., Evans, L., Wales, K., Leonard, C., & Lane, A. E. (2023). Measurement of self-regulation in preschool and elementary children: A scoping review. *Physical & Occupational Therapy in Pediatrics*, *43*(4), 403–429. https://doi.org/10.1080/01942638.2022.2158055

Pollock, D., Davies, E. L., Peters, M. D. J., Tricco, A. C., Alexander, L., McInerney, P., & Munn, Z. (2021). Undertaking a scoping review: A practical guide for nursing and midwifery students, clinicians, researchers, and academics. *Journal of Advanced Nursing*, *77*(4), 2102–2113. https://doi.org/10.1111/jan.14743

Polya, G. (1945). *How we solve it. A new aspect of mathematical method*. Princeton.

Popham, W. J. (2017). *Classroom assessment: What teachers need to know* (8th ed.). Pearson.

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

Qiao, H., & Huang, J. (2023). Investigating children's mathematical learning during block play: A video observation research. *Early Child Development and Care*, *193*(13–14), 1445–1458. https://doi.org/10.1080/03004430.2023.2256992

Ray, W. S. (1955). Complex tasks for use in human problem-solving research. *Psychological Bulletin*, *52*(2), 134–149. https://doi.org/10.1037/h0044763

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*, 1380178. https://doi.org/10.3389/fpsyg.2024.1380178

Rogers, J. (2004). Autonomy and mathematical problem-solving: The early years. *Education*, *32*, 24–31. https://doi.org/10.1080/03004270485200291

Ros, R., Bjarnason, E., & Runeson, P. (2017). A machine learning approach for semi-automated search and selection in literature studies. Proceedings of the 21st international conference on evaluation and assessment in software engineering, Karlskrona, Sweden https://doi.org/10.1145/3084226.3084243

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

Said-Metwaly, S., Noortgate, W. V., & Kyndt, E. (2017). Methodological issues in measuring creativity: A systematic literature review. *Creativity. Theories – Research – Applications*, *4*(2), 276. https://doi.org/10.1515/ctra-2017-0014

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.

Seitz, H., & Bartholomew, C. (2008). Powerful portfolios for young children. *Early Childhood Education Journal*, *36*(1), 63–68. https://doi.org/10.1007/s10643-008-0242-7

*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

Silva, C., Sousa-Gomes, V., Fávero, M., Oliveira-Lopes, S., Merendeiro, C. S., Oliveira, J., & Moreira, D. (2022). Assessment of preschool-age executive functions: A systematic review. *Clinical Psychology & Psychotherapy*, *29*(4), 1374–1391. https://doi.org/10.1002/cpp.2718

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

Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. Estes (Ed.), *Handbook of learning & cognitive processes: V. Human information* (pp. 271–295). Lawrence Erlbaum.

Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. The National Academies 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.

Suherman, S., & Vidákovich, T. (2022). Assessment of mathematical creative thinking: A systematic review. *Thinking Skills and Creativity*, *44*, 101019. https://doi.org/10.1016/j.tsc.2022.101019

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sundqvist, P., & Nilsson, T. (2018). Technology education in preschool: Providing opportunities for children to use artifacts and to create. *International Journal of Technology and Design Education*, *28*(1), 29–51. https://doi.org/10.1007/s10798-016-9375-y

Suto, I. (2013). 21st century skills: Ancient, ubiquitous, enigmatic? *Research Matters: A Cambridge Assessment Publication*, *15*, 100282.

Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 103798. https://doi.org/10.1016/j.compedu.2019.103798

Tank, K. M., Moore, T. J., Dorie, B. L., Gajdzik, E., Terri Sanger, M., Rynearson, A. M., & Mann, E. F. (2018). Engineering in early elementary classrooms through the integration of high-quality literature, design, and stem+c content. In L. English & T. Moore (Eds.), *Early engineering learning* (pp. 175–201). Springer. https://doi.org/10.1007/978-981-10-8621-2_9

Torrance, E. P. (1974). *Torrance tests of creative thinking: Directions manual and scoring guide*. Personnel Press Lexington.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/m18-0850

*Tsamir, P., Tirosh, D., Tabach, M., & Levenson, E. (2010). Multiple solution methods and multiple outcom: Is it a task for kindergarten children? *Educational Studies in Mathematics*, *73*(3), 217–231. https://doi.org/10.1007/s10649-009-9215-z

Tunnicliffe, S. D. (2022). Play and STEM foundations in the earliest years. In S. D. Tunnicliffe & T. J. Kennedy (Eds.), *Play and STEM education in the early years: International policies and practices* (pp. 39–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-99830-1_2

Ukobizaba, F., Nizeyimana, G., & Mukuka, A. (2021). Assessment strategies for enhancing students' mathematical problem-solving skills: A review of literature. *EURASIA Journal of Mathematics, Science and Technology Education*, *17*(3), em1945. https://doi.org/10.29333/ejmste/9728

*Unal, M., & Aral, N. (2014a). Development of the problem solving scale in science education (PSSSE): The reliability and validity study. *Egitim Ve Bilim [Education and Science]*, *39*(176), 267–278. https://doi.org/10.15390/EB.2014.3585

*Unal, M., & Aral, N. (2014b). An investigation on the effects of experiment based education program on six years olds' problem solving skills. *Egitim Ve Bilim [Education and Science]*, *39*(176), 279–291. https://doi.org/10.15390/EB.2014.3592

*Unal, M., & Saglam, M. (2018). Examination of the effect of the GEMS program on problem solving and science process skills of 6 years old children. *European Journal of Educational Research*, *7*(3), 567–581. https://doi.org/10.12973/eu-jer.7.3.567

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

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, *3*(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7

*Van Elsen, J., Catrysse, L., & De Maeyer, S. (2024). The effect of interactive picturebook reading on problem-solving skills in preschool: A quasi-experiment. *Early Childhood Education Journal*, *52*(7), 1471–1485. https://doi.org/10.1007/s10643-023-01542-3

Varghese, V. V. V., & Renumol, V. G. (2021). Assessment methods and interventions to develop computational thinking: A literature review. 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India https://doi.org/10.1109/ICITIIT51526.2021.9399606

*Vessonen, T., Hellstrand, H., Aunio, P., & Laine, A. (2024). Individual differences in mathematical problem-solving skills among 3- to 5-year-old preschoolers. *International Journal of Early Childhood*, *56*, 339–357. https://doi.org/10.1007/s13158-023-00361-2

Wagner, G., Lukyanenko, R., & Paré, G. (2022). Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology*, *37*(2), 209–226. https://doi.org/10.1177/02683962211048201

Wan, Z. H., Jiang, Y., & Zhan, Y. (2021). STEM education in early childhood: A review of empirical studies. *Early Education and Development*, *32*(7), 940–962. https://doi.org/10.1080/10409289.2020.1814986

Wang, G., & Chen, S. (2025). Integrating STEM principles into kindergarten science education. *International Journal of Knowledge Management (IJKM)*, *21*(1), 1–17. https://doi.org/10.4018/IJKM.383964

Wang, Z., Nayfeh, T., Tetzlaff, J., O'Blenis, P., & Murad, M. H. (2020). Error rates of human reviewers during abstract screening in systematic reviews. *PLoS One*, *15*(1), e0227742. https://doi.org/10.1371/journal.pone.0227742

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ward, S. D. (1993). Developmental versus academic mathematics education: Effects on problem solving performance and attitudes toward mathematics in kindergarten children. [Unpublished doctoral dissertation]. Vanderbilt University.

Wicaksono, A. G. C., & Korom, E. (2022). Review of problem-solving measurement: An assessment developed in the Indonesian context. *Participatory Educational Research*, *9*(1), 116–136. https://doi.org/10.17275/per.22.7.9.1

How to cite this article: Van Elsen, J., De Maeyer, S., & Torbeyns, J. (2026). Methods for assessing problem-solving competencies in preschoolers: A scoping review. *Review of Education*, *14*, e70181. https://doi.org/10.1002/rev3.70181

20496613, 2026, 2, Downloaded from https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.70181 by Joris Van Elsen - Universiteit Antwerpen , Wiley Online Library on [10/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License