The Effect of Interactive Picturebook Reading on Problem-Solving Skills in Preschool: A Quasi-Experiment

Joris Van Elsen[1], Leen Catrysse[2], Sven De Maeyer[1]

[1] *EduBROn, University of Antwerp (Belgium)*

[2] *Department of Online Learning and Instruction, Faculty of Educational Sciences, Open Universiteit (Netherlands)*

KEY WORDS

Problem solving
Interactive picturebook reading
Preschool education
Quasi-experiment

This manuscript has been accepted for publication in Early Childhood Education Journal on June 23, 2023. The published article is available online at: https://doi.org/10.1007/s10643-023-01542-3

ABSTRACT

Interactive picturebook reading is an easily accessible and enjoyable activity that provides preschoolers with opportunities to practice problem-solving skills. Increasingly, preschool age is considered the ideal time to foster problem-solving skills. This paper examines the effect of interactive picturebook reading on the problem-solving skills of preschool children. A total of 130 preschool children between 4 and 6 years old participated in a pretest-posttest quasi-experiment with a non-equivalent control group. During five interactive reading sessions in small groups, preschoolers were invited to discuss possible solutions to the problems the book characters encounter. Problem-solving skills were measured with two Pictorial Multiple-Solutions Tasks. Multilevel analysis indicates that interactive picturebook reading can improve children's flexibility, fluency and originality in coming up with possible solutions to a problem. In addition, we found that the picture that was used and the emotional reactions of the children to the test situation influenced the test scores. The results show that giving children the opportunity to discuss possible solutions to problems presented in picturebooks is an effective way to work on problem-solving skills in preschool. Further research is needed to explore the optimal conditions for effective interactive picturebook reading to develop preschoolers' problem-solving skills and the extent to which the skills demonstrated during interactive picturebook reading can be transferred to real-life problem solving.

Interactive picturebook reading offers plenty of opportunities to practice problem-solving skills in preschool (Gosen et al., 2015). Picturebooks are packed with problems that need to be solved. Fortunately, preschoolers are eager to help the book characters to solve their problems by providing possible solutions (Gosen et al., 2015). Problem solving is an important 21st century skill and is required to live and learn in our ever-changing world (Csapo & Funke, 2017). Increasingly, researchers acknowledge that young children are able to solve complex problems and that preschool age is an ideal time to foster the development of problem-solving skills (e.g., Anggoro et al., 2021). Young children have a seemingly insatiable curiosity and urge to explore the world around them (Fusaro & Smith, 2018). By tapping into this need for exploration, preschool teachers can maintain, stimulate and strengthen children's problem-solving skills. In this study, we investigate the effect of interactive reading from picturebooks on preschoolers problem-solving skills.

Interactive Picturebook Reading

The importance of picturebook reading in preschool is recognized across continents and disciplines. The effect of picturebook reading on young children's language development is well documented (Mol et al., 2009). In recent years, there has been growing scientific interest in the effect of picturebook reading on other domains such as mathematics (Op 't Eynde et al., 2022; Zhang et al., 2023) and social skills or attitudes (e.g., Kemple et al., 2016). However, picturebook reading is not a panacea in itself. The 'magic' occurs when you let children interact with the story, the teacher and each other (Meyer et al., 1994). Shared reading becomes interactive when input from the children is asked before, during or after reading (Massey, 2004). Interactive picturebook reading encompasses a wide range of practices and approaches where children are engaged during shared reading. Teachers can elicit interaction by asking questions, provoking discussions or inviting children to make predictions. Interactive reading can differ in terms of the moment at which the interactions

take place (before, during and/or after reading aloud) and the type and desired level of the interactions (Gosen et al., 2015).

The added value of interaction during picturebook reading is well investigated (e.g., Merga, 2017; Mol et al., 2009). Interactive reading contributes to the development of complex thinking and encourages children to co-construct meaning through dialogue (Wiseman, 2011). Gosen et al. (2015) found that talk about problems during picturebook reading offered preschoolers opportunities to practice problem-solving skills. Most research on interactive picturebook reading focusses on the interaction between the adult and the children (van Kleeck, 2003). Morrow and Smith (1990) pointed out that stimulating interaction between children can enhance the effect of interactive picturebook reading. Interaction among children during shared reading allows them to try things out, think through solutions, and reflect on their ideas (Pantaleo, 2007). Interactive picturebook reading provides children with opportunities to identify problems, to think about possible solutions and to think about their consequences (Fettig et al., 2016). In addition, picturebooks enable children to think creatively about situations and problems they have never experienced before (Gosen et al., 2015). Despite the popularity of shared reading interventions with picturebook in early childhood education and research (Fitton et al., 2018), the effect of interactive picturebook reading on preschoolers' problem-solving skills has not yet been investigated. The current study extends the research on the effects of interactive picturebook reading with problem-solving, an important, transversal skill needed for many subject areas and strongly related to self-regulated learning and lifelong learning (Baars et al., 2017). Moreover, insights on the effect of interactive picturebook reading on problem-solving skills can provide practitioners with evidence about the potential of picturebooks already present in their classrooms when working around problem solving.

Problem Solving

Problem solving is the act of solving unknown problems where no obvious solution is at hand (Mayer, 1994). It can be defined as "an individual's capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious" (OECD, 2013, p. 122). Problem-solving encompasses a wide range of cognitive and non-cognitive activities that involve seeking a solution to a given problem (Green & Gilhooly, 2012). To get a better grasp on the complexity of problem solving, many scholars have tried to identify different steps in the problem-solving process. The PISA framework on problem solving distinguishes between (1) Exploring and understanding (2) Representing and formulating (3) Planning and executing, and (4) Monitoring and reflecting (OECD, 2013).

The current study focuses on the second step of the problem-solving process as proposed by the PISA framework and more specifically on formulating possible solutions to a given problem. Osborn (1953) considers this piling up of alternative hypothesis as "probably the one indispensable part of any problem-solving project" (p. 125). This step is strongly related to creative thinking or divergent production (Guilford, 1967). Divergent production is conceptualized as the generation of various logical possibilities or, more generic, the ability of coming up with multiple solutions to a problem. Gosen et al. (2015) demonstrated that children between 4 and 6 are capable of hypothesizing during shared reading from picturebooks. Central to measuring creativity in problem solving are the concepts of fluency, flexibility and originality (Said-Metwaly et al., 2017). Fluency refers to the number of unique solutions one can think of. Flexibility refers to the different categories of these solutions (methods, strategies, principles or perspectives). Originality refers to the uniqueness of the solutions.

The Present Study

Building on the extensive literature on problem solving and interactive picturebook reading interventions, the current study aims to examine the effect of interactive reading from picturebooks in the classroom on the extent to which preschool children can come up with different solutions to a problem. The resulting research question can be formulated as:

> What is the effect of interactive picturebook reading in the classroom on preschoolers' ability to come up with different solutions for a visually presented problem?

Method

In order to answer the research question, a pretest-posttest quasi-experimental study with a non-equivalent control group was set up (Cohen et al., 2018). The study was approved by the Ethical Advisory Committee for Social and Human Sciences (EASHW) of the University of Antwerp on 12 January 2021 (file SHW_20_112).

Participants

Preschool Children

A total of 130 preschool children from nine preschool classes in three Flemish (Belgium) schools participated in this study. The schools were located in a suburban area and had a predominantly white student population. Classes were selected based on willingness and availability of the teacher (convenience sampling). In the sample, there were 70 girls (53.85 %) and 60 boys (46.15 %). The mean age of the participants was 5.57 year (*Min* = 4.17, *Max* = 6.67). Eight of them spoke a different language at home than the language of the picturebooks and the tests. Around 70% of participating children had at least one parent with a bachelor's degree or equivalent. Only children for whom parental consent was obtained, took part in the experiment.

Teachers

The participating teachers were between 28 and 54 years old and had 6 to 32 years of experience in preschool education. One teacher co-taught two classes with a colleague and one participant was a care teacher. The smallest class had 18 preschoolers, the largest had 22. Six classes were assigned to the treatment group and three classes to the control group.

Measures and Materials

Picturebooks

A different picturebook was used for each session of the intervention. Using the same set of picturebooks in all the groups was less appropriate for this study. Participants were not allowed to know the picturebooks in advance since preschoolers who already knew the story might be inhibited in coming up with alternative solutions. To help teachers select appropriate picturebooks for the interventions, a list of popular and readily available picturebooks was provided. The list was composed based on the list of picturebooks provided in the doctoral dissertation of Gosen (2012), library catalogs and catalogs of Flemish and Dutch publishers of picturebooks. The list was completed in consultation with a children's librarian at a public library, an owner of a specialized children's bookshop and preschool teachers. This resulted in a longlist of 30 picturebooks. Picturebooks that were hard to find or no longer available, stories in which the problem was only implicitly or vaguely introduced and picturebooks that didn’t meet the quality criteria concerning print, structure, pictures and wordings, were removed from the list. Finally, 12 picturebooks were retained. Teachers were allowed to choose other picturebooks for the intervention provided that they met the pre-established criteria included in the logbooks. The selected picturebooks had to be appropriate for preschoolers in terms of difficulty (Nicolopoulou et al., 2023) and storyline and they had to include problems to be solved by the main characters. All picturebooks were written in Dutch. Table 1 provides an overview of the picturebooks used during the interventions and the number of interventions in which the picturebooks were used. During most sessions (89%) a picturebook from the proposed list was used.

Table 1. Overview of the picturebooks used during the interactive reading sessions

Title [English]	Author(s)	Year	Sessions (*N*)
**Garage Gust [Gu's garage]*	Timmers, L.	2015	15
**Een gat in mijn emmer [A hole in my bucket]*	Schubert, I. & Schubert, D.	1998	13
**Diepzeedokter Diederik [Deep-sea doctor Derek]*	Timmers, L.	2017	11
**Agent en Boef [Cop and Robber]*	Veldkamp, T. & de Boer, K.	2008	9
**De prinses met de lange haren [The princess with the long hair]*	Van Haeringen, A.	2015	9
**Twee vechtende eekhoorntjes [The squirrels who squabbled]*	Bright, R. & Field, J.	2017	5
Rikkie heeft een plan [Rikkie has plan]	Van Genechten, G.	2015	4
**Hoe verstop je een giraffe? [Have you seen my giraffe?]*	Robinson, M. & Powell, C.	2018	3
Happende haaien [Yummy yummy! Food for my tummy!]	Lloyd, S. & Tickle, J.	2004	2
Kikker vindt een schat [Frog finds a treasure]	Velthuijs, M.	2017	1
De sleutel [The key]	Flas, I. & Masson, A.	2012	1

Note. Picturebooks marked with an asterisk were included in the proposed list.

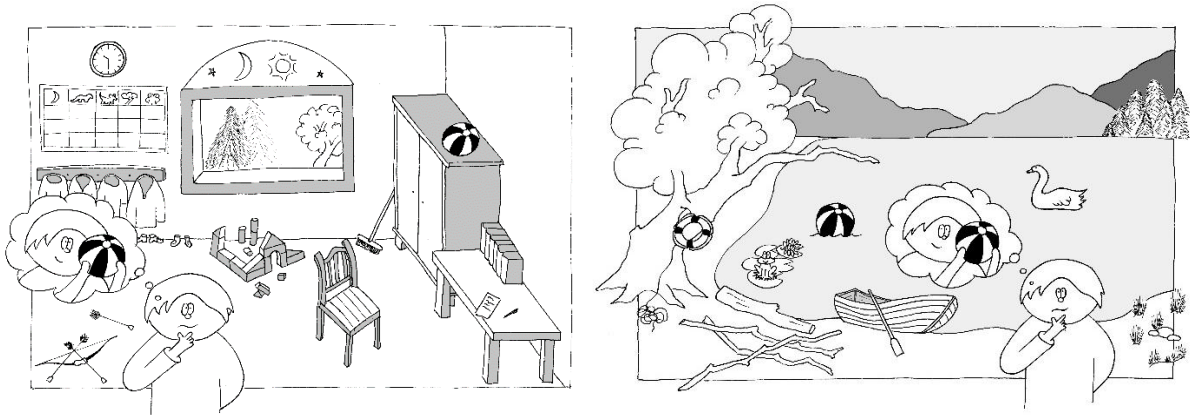
Short Questionnaire on Child Characteristics

Along with the information and consent letter, parents received a short questionnaire asking for gender, age (year and month of birth) and language spoken at home. Previous studies indicate that these characteristics are correlated with problem-solving skills and may influence the effect of a reading intervention (Casey et al., 2008; van den Heuvel-Panhuizen et al., 2016).

Pictorial Multiple-Solutions Tasks

Problem-solving skills were measured through Pictorial Multiple-Solutions Tasks (PMST). In PMSTs, preschool children are asked to come up with as many solutions as possible to a visual presented problem (Leikin, 2012). Two PMSTs were designed for the current study (Fig. 1). The problem situation in both pictures is similar: an androgynous character cannot reach the ball. In one picture the ball is on a cupboard, in the other picture the ball is in the pond. Both pictures show a similar number of attributes that can be used or combined to solve the problem. These attributes are linked to problem-solving strategies such as tossing, climbing or pulling or pushing. The children were asked to suggest as many solutions as they could to solve the presented problem, within a timeframe of approximately three minutes. The teachers were allowed to encourage the children with prescribed phrases. Teachers were also allowed to support children with language difficulties by giving them more time to formulate solutions and by motivating them to point out objects in the pictures and use gestures to clarify what actions could be taken.

Fig. 1 Pictorial Multiple-Solutions Tasks



PMST 1: Ball on cupboard **PMST 2**: Ball in pond

Both PMSTs were subjected to a pilot test beforehand. Four independent researchers scored two videotaped test takings in which two children each took one of the tests. The inter-rater reliability (ICC) and

95% confidence intervals were calculated using a single-rater, absolute-agreement, two-way random-effects model (Koo & Li, 2016). The ICC for the test with the ball on the cupboard was 0.89 (95 % CI [0.71, 0.98]), which is considered good (Cicchetti, 1994). The ICC for the test with the ball in the pond was 0.69 (95 % CI [0.27, 0.96]). This is slightly lower than the other test, but still good. Next, the video of the test with the ball on the cupboard was scored by four participating teachers. They all noted down exactly the same solutions which resulted in perfect agreement (ICC = 1).

Scorebooks

A scorebook was developed for each test. The most obvious solutions were included in the scorebook and space was provided to note less obvious solutions. In addition to recording the responses of the children, the teachers were asked to indicate in the scorebooks whether a child was nervous or shy, showed little interest, or was easily distracted during the test. In previous studies, researchers noticed that these factors can influence the test-taking and test results (Rathé et al., 2016; Tsamir et al., 2010).

Logbooks

Teachers were asked to complete a research-driven logbook after each interactive reading sessions. Research-driven logbooks provide an opportunity to gather background information on the intervention. The logbook included questions about the title of the book, the duration of the reading sessions, the children who participated in the session and statements to evaluate the quality of the session. Logbooks and a debriefing after the experiment were used to assess the extent to which the treatment was carried out in accordance with the protocol (Capin et al., 2018).

Procedures

Intervention

The intervention was carried out by the preschool teachers. Having teachers read to their own children increases ecological validity. Teachers were asked to interactively read from picturebooks in small groups of 5 to 7 children (Morrow & Smith, 1990; van Elsäcker & Verhoeven, 1997) for five consecutive days. To ensure that children in all classes received an equivalent intervention, a protocol was developed for the interactive reading sessions. The protocol was provided in text and verbally explained during an individual online meeting. The protocol included guidelines for the choice of picturebooks, the organization of the reading session, the place and duration of the sessions, envisaged types of interaction, key issues, and examples of questions to elicit interaction. These questions entailed open-ended wh-questions (e.g., “What (else) could you do?”) (Walsh & Hodge, 2018), questions for clarification (e.g., "How exactly would you do that?", “You said you would ask for help. Who do you think could help?”) and evaluative questions (e.g., “Is this a good idea? What might happen if you do?”, “What do you think is the best of all suggested solutions?”). The protocol also prescribed how to engage the children in the story by asking questions, responding to their comments, engaging with them in dialogue or encouraging interactions among the children (Towson et al., 2017).

After each intervention, teachers were asked to fill in the logbook by providing information and perceptions about the reading session (see above). A total of 73 reading sessions were conducted. On average, the reading sessions lasted for 20 minutes (*Min* = 10, *Max* = 30, *SD* = 5.70). Group sizes ranged from four to eight children (*M* = 6.12, *SD* = 0.98). One teacher read to the entire class (17 to 19 children) and one teacher started the first session for the entire class and continued in groups of four to five children. The reading sessions occurred in another (class)room or a quieter corner of the classroom. In nearly all cases, a second teacher (care teacher or co-teacher) kept an eye on the other children.

Analysis of the logbooks confirmed that the reading sessions were carried out as planned (adherence and dosage). Children who participated in less than three reading sessions were excluded from the analysis because of lack of exposure (Sanetti et al., 2021). Due to the restrictions caused by the COVID-19 measures at the time of study, it was not permitted to observe interventions.

Tests

Both pretest and posttest were administered in the children's own classroom or a separate room in the school in which the children feel comfortable. Rathé et al. (2016) found that some preschoolers are shy when an unknown researcher administers a test individually. By having the tests administered by their own teachers, we tried to limit the impact of shyness on the test results. The children were asked to suggest as many solutions as they could to solve the presented problem, within a timeframe of approximately three minutes. The teachers were allowed to encourage the children with prescribed phrases. A protocol was developed to ensure that all teachers administered the test in a uniform way. The protocol was included in the scorebooks, explained verbally during an individual online meeting and illustrated with an instruction and practice video (Van Elsen, 2020).

For every child, a fluency score was calculated by counting the number of different formulated solutions (Leikin, 2012). Subsequently, all solutions were classified into categories related to an action (e.g., throwing, climbing, pulling) or to specific characteristics of the solution (e.g., use of materials not shown, inventive solutions, combinations of actions). During the classification phase, new categories were created for unforeseen solution strategies. A flexibility score was calculated for every child based on the number of categories into which at least one answer was classified.[1] Table 2 provides an overview of the number of solutions the children formulated for each category and test. The numbers include solutions from both the pretest and the posttest and the treatment and control group. Solutions that did not fit well in one of the categories were given their own category (e.g., "Looking in the books for a solution", "Growing up", and "Buying a new ball"). The test with the ball in the pond produced slightly more different solutions (*N* = 103) than the test with the ball on the cupboard (*N* = 98).

Table 2 Overview of the categories and the number of different solutions per category per test

	Cupboard	Pond
Climbing	9	1
Pulling or pushing	5	4
Tossing	13	7
Shooting (cupboard) / boating (pond)	2	3
Stacking (cupboard) / swimming (pond)	12	3
Crafting	3	16
Use of material not depicted	10	6
Without material (jumping, blowing ...)	7	4
Ask for help	2	1
Fantasy	12	23
Combining objects and/or actions	21	32
Own category	2	3
Total	98	103

Note. The number of different solutions is based on the answers from the pretest and the posttest together.

To determine the originality of a solution, the number of children who gave a similar answer was considered. Any solution that was given by up to 10 % of the children is considered original and scores 2 points. Here, we assume that if only one in ten children can come up with a particular solution, it is a non-obvious solution. Any solution that was given by 10 % to 25 % of the participants scores 1 point. Solutions given by more than a quarter of the children do not yield points. When more than one in four children came up with a particular solution, we assume that the solution can be fairly easily derived from the picture. For the calculation of the originality score, both the answers from the pretests (*N* = 130) and the posttests (*N* = 125) were used as a reference to define its originality. The originality score for every child is the sum of the originality-scores of every solution the child came up with.

[1] Note that the calculation for flexibility and originality differ from the formula used by Leikin (2012). The picture Leikin used contained fewer attributes and suggested fewer possibilities than the pictures in the current study. This necessitated adjustments to the calculation of the flexibility and originality scores.

Analysis

Multilevel analysis was used to investigate the effect of the intervention on the children's fluency, flexibility and originality scores, taking into account child characteristics and variables related to the test taking. Multilevel modelling is appropriate because the pretest and posttest scores (level 1) are 'nested' in children (level 2) and children are nested in classes (level 3) (Hox et al., 2018).

Building the model was done in two phases. First, a basic model was constructed. The basic model is a model without covariates in the fixed part, except for the necessary dependent variables indicating the test (pretest or posttest), condition (treatment or control) and their interaction (treatment:condition). Stepwise, more complexity was added to the random part of the basic model. Successively, the child level and the classlevel are added. Finally, we allowed the basic model to estimate differentiated child-level variance for the pretest and the posttest. After all, we assume that the treatment does not only affect the test scores, but also the variance, since the true variance in posttest scores is in fact a function of the true variance in the pretest score and the true variance of the learning gains from the intervention (Zimmerman & Williams, 1982).

In the second phase, covariates were added to the model blockwise. The first block of covariates relates to the lowest level: the level of the test. We consider that the two pictures that were used for the pretest and the posttest differ from each other and that children's reactions to the test can affect the results. Therefore, we included the picture (cupboard or pond) and dummy-variables indicating the child was shy, easily distracted or showed little interest in the fixed part of the model (Model 2). The second block of covariates relates to the test-independent child level characteristics. In this step, covariates were added for gender, whether the child spoke a different language at home than the language of the test (dummy variables) and age (standardised, decimal) (Model 3). Finally, two interaction terms were added as fixed effects to the model (Jaccard & Turrisi, 2003).

To select the most parsimonious model, the models are compared using Akaike's Information Criterion (AIC), conditional R^2 and marginal R^2. The AIC can be interpreted as the deviation between the predicted values and the observed values, corrected for the number of independent variables in that model (Akaike, 1974). The lower the AIC, the better the model approaches reality in comparison to the other models. The conditional R^2 is the proportion of the variance explained by the full model. The marginal R^2 provides information about the proportion of the total variance explained by the fixed effects (Nakagawa et al., 2017). A marginal R^2 of .26 or more is considered substantial, an R^2 around .13 is considered moderate and an R^2 of .02 or lower is considered weak (Cohen, 1988).

The assumptions for multivariate regression of the best performing models were checked (Hox et al., 2018). After removing outliers and re-estimating the models, we calculated differences between group means (contrasts). Finally we calculated Cohen's *d* based on the second formula provided by Morris (2008) to estimate the effect size for fluency and flexibility.

Multilevel analyses were performed in R 4.2.1 (R Core Team, 2022) with the package lme4 (v1.1-30; Bates et al., 2015) and blme (v1.0-5; Chung et al., 2013). For testing models, the packages lmerTest (v3.1-3; Kuznetsova et al., 2017) and performance (v0.9.2; Lüdecke et al., 2021) were used. Estimated marginal means were computed with emmeans (v1.8.1-1; Lenth, 2022). Figures were created with ggplot2 (v3.3.6; Wickham, 2016).

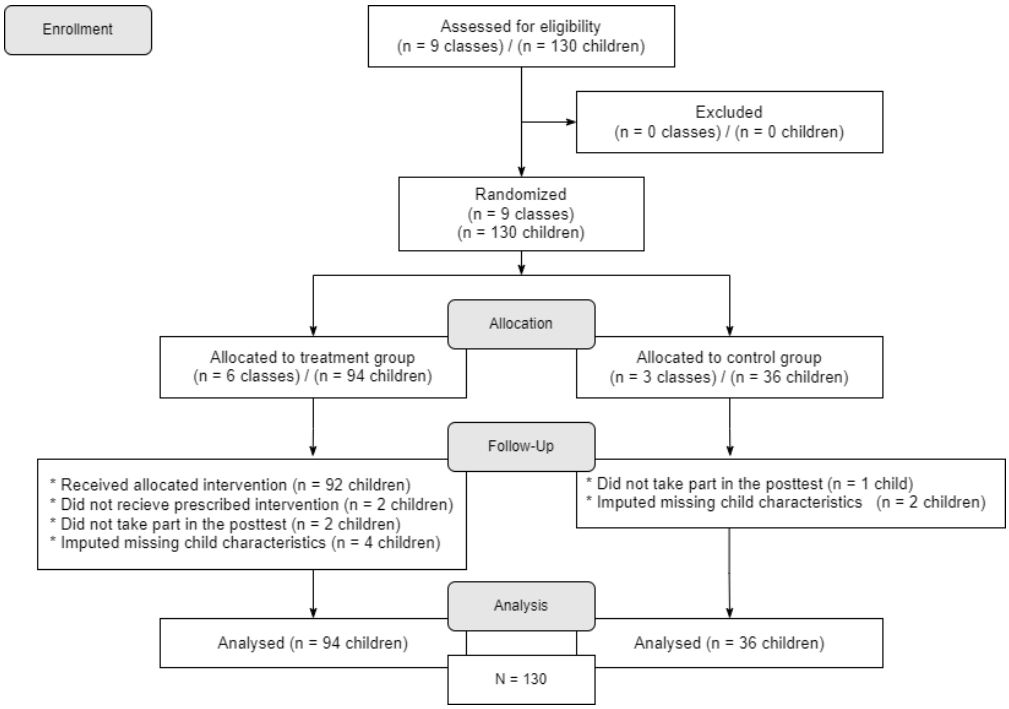
Results

Missing Data

Three children did not take the posttest. Two children attended less than three reading sessions. For these cases, the posttest results were removed from the dataset. No child characteristics were provided for three children and the age of two other children was missing. To avoid the loss of valuable data, missing ages were imputed by the median age. If no information about the home language was provided, it was assumed that the child in question did not speak any other language at home. Fig. 2 provides an overview of the number

of participants through the different phases of the experiment (enrolment, allocation, follow-up and analysis).

Fig. 2 CONSORT flow diagram of the present quasi-experiment (Schulz et al., 2010)



Descriptive Statistics

Descriptive statistics for each factor, test and group are summarized in Table 3. The figures reveal large differences between children. Four children formulated only one solution during the pretest, whereas another child was able to suggest 17 solutions. Similarly, some children scored extreme low for flexibility and originality, whereas some others were able to suggest solutions from eight different categories and up to twelve original solutions resulting in an originality score of 24. It should be noted that fluency, flexibility and originality are strongly correlated. The correlation between fluency and flexibility for the pretest (r = .80, p < .001) and between fluency and originality (r = .91, p < .001) is very strong. The correlation between flexibility and originality is somewhat smaller, but, according to the rules of thumb of J. Cohen (1988), the correlation remains strong (r = .63, p < .001).

Table 3 Descriptive statistics per group, test and picture

		Fluency				Flexibility				Originality			
	N	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
Pretest	130	5.50	2.96	1	17	3.70	1.65	1	8	3.63	4.14	0	24
Treatment group	94	5.49	3.16	1	17	3.64	1.73	1	8	3.73	4.38	0	24
Cupboard	66	5.59	3.26	1	17	3.44	1.71	1	8	4.05	4.79	0	24
Pond	28	5.25	2.93	1	13	4.11	1.73	1	8	3.00	3.17	0	10
Control group	36	5.53	2.40	1	12	3.86	1.40	1	7	3.36	3.48	0	14
Cupboard	20	6.05	2.80	2	12	3.70	1.56	1	7	4.65	4.15	0	14
Pond	16	4.88	1.63	1	7	4.06	1.18	1	5	1.75	1.24	0	4
Posttest	125	6.19	2.44	1	13	4.77	1.65	1	9	3.69	3.08	0	13
Treatment group	90	6.54	2.40	2	13	4.97	1.56	1	9	4.16	3.18	0	13
Cupboard	28	5.36	1.85	2	10	3.96	1.35	1	7	2.96	2.43	0	9
Pond	62	7.08	2.45	2	13	5.42	1.44	2	9	4.69	3.34	0	13
Control group	35	5.29	2.33	1	11	4.26	1.77	1	8	2.49	2.49	0	10
Cupboard	15	5.13	2.17	2	10	3.53	1.46	1	6	2.27	2.22	0	7
Pond	20	5.40	2.50	1	11	4.80	1.82	1	8	2.65	2.72	0	10

Although the differences between scores remain large in the posttest, the number of children scoring extremely low is strongly reduced. During the pretest, 16 children (12.8 %) suggested only one or two solutions. During posttest, only 5 (4.0 %) children did so. The same applies for the flexibility scores. During the pretest, 27 children (21.6 %) formulated solutions from only one or two categories. During the posttest, the number of children with such a low score was reduced to 11 (8.8 %). Children who formulated few solutions or only solutions that could be derived directly from the picture did not score any points for originality. However, the number of children who did not suggest at least one original solution decreased from the pretest ($N = 26$; 20.8 %) to the posttest ($N = 17$; 13.6 %). Fig. 3 presents the differences in distributions between the pretest and the posttest, separately for the treatment group and the control group. The distribution of the flexibility scores for the treatment group shifts to the right. To a lesser extent, the same applies for fluency and originality. Concerning the control group, there are no obvious differences between the pretest and the posttest.

Fig. 3 Distribution of the fluency, flexibility and originality scores for pretest and posttest

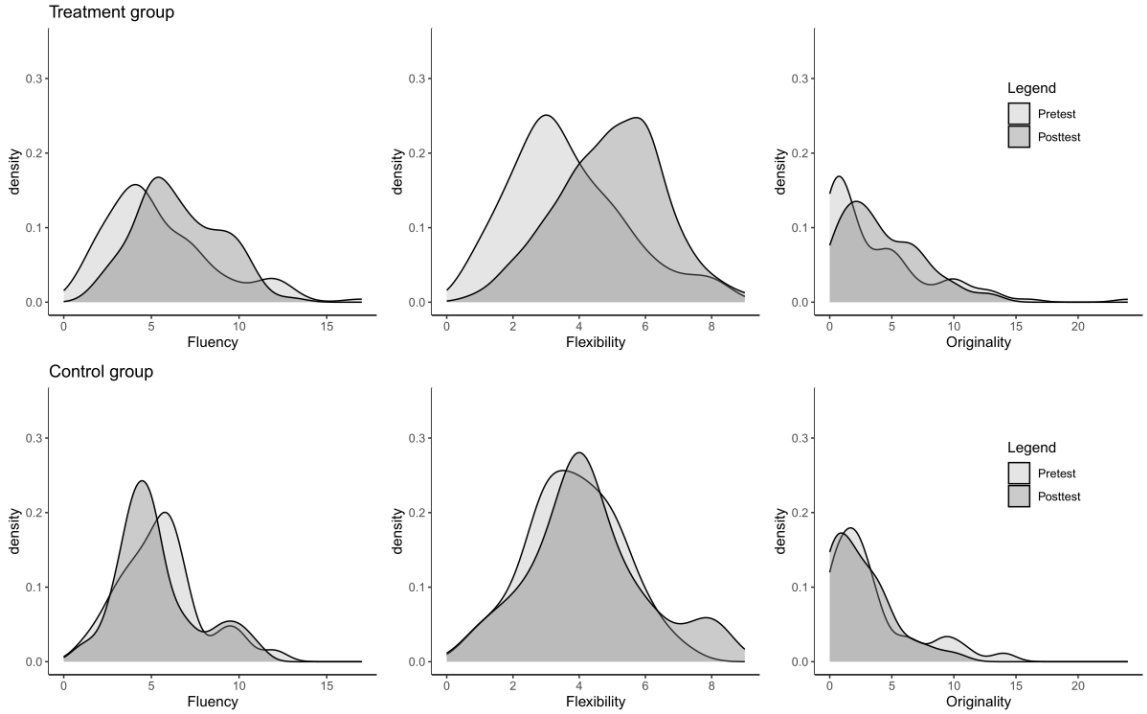


Table 4 provides an overview of the number and percentage (between brackets) of children whose teacher indicated that they were shy, distracted or showed a lack of interest during the test taking. The proportion of children whose teacher indicated that they were shy was smaller in the posttest (7.20 %) than in the pretest (12.31 %). The same applies for distraction (from 7.69 % to 3.20 %) and to a lesser extent to lack of interest (from 4.62 % to 4.00 %).

Table 4 Number (percentage) of children whose teacher indicated that they were shy, distracted or showed a lack of interest during the test taking

	N		Shy		Distracted		Lack of interest	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Total	130	125	16 (12.31)	9 (7.20)	10 (7.69)	4 (3.20)	6 (4.62)	5 (4.00)
Treatment group	94	90	13 (13.83)	7 (7.78)	7 (7.45)	2 (2.22)	5 (5.32)	4 (4.44)
Cupboard	66	28	8 (12.12)	1 (3.57)	7 (10.61)	0 (0.00)	4 (6.06)	0 (0.00)
Pond	28	62	5 (17.86)	6 (9.68)	0 (0.00)	2 (3.23)	1 (3.57)	4 (6.45)
Control group	36	35	3 (8.33)	2 (5.71)	3 (8.33)	2 (5.71)	1 (2.78)	1 (2.86)
Cupboard	20	15	1 (5.00)	1 (6.67)	0 (0.00)	1 (6.67)	0 (0.00)	0 (0.00)
Pond	16	20	2 (12.50)	1 (5.00)	3 (18.75)	1 (5.00)	1 (6.25)	1 (5.00)

Statistical Models

As illustrated by Fig. 3, the originality scores are far from normally distributed. The distribution of all four groups is strongly right skewed (skewness > 1) and on the pretest, 42.31 % of the children scored 0 or 1 for originality. On the other hand, some children scored extremely high. This was to be expected given the way the originality score was calculated. We can presume that children that only suggest a limited number of possible solutions, only come up with more obvious solutions. Consequently, the originality scores tend to a Poisson distribution. Therefore, we use a Generalised Mixed-Effects Model (GLMM) for originality (Cameron & Trivedi, 1998).

Concerning fluency and originality, the basic models with varying variances between the pretest and the posttest on the child level perform best (Table 5). The AIC of these model is significantly lower than the AIC from the more sparse models ($p < 0.05$). Concerning Flexibility, the results indicate that allowing the model to estimate varying variance on the child level, does not improve the model.

Table 5 Model comparison of the basic models

Fluency	*npar*	*AIC*	*-2LL*	*Chi²*	*DF*	*Pr(>Chi²)*
Model 1a: Child level	6	681.18	669.18			
Model 1b: Child and class level	7	658.01	644.01	25.17	1	<.001
Model 1c: Varying variance on child level	**9**	**652.00**	**634.00**	**10.02**	2	**.007**
Flexibility						
Model 1a: Child level	6	683.89	671.89			
Model 1b: Child and class level	**7**	**676.40**	**662.40**	**9.49**	**1**	**.002**
Model 1c: Varying variance on child level	9	680.40	662.40	0.00	2	.998
Originality						
Model 1a: Child level	5	1199.40	1189.40			
Model 1b: Child and class level	6	1181.80	1169.80	19.55	1	<.001
Model 1c: Varying variance on child level	**8**	**1154.20**	**1138.20**	**31.62**	2	**<.001**

Note. npar = number of parameters in the model; AIC = Akaike Information Criteria, -2LL = deviance; Chi² = Δ-2LL; DF = degrees of freedom for the likelihood ratio test; Pr(>Chi²) = p-value of the Chi². Bold = selected models.

Starting from the basic model, covariates were added to account for test-related conditions, child characteristics and differential intervention effects. Model comparisons indicates that adding covariates related to the test level (the picture used and reactions of the children to the test), improved all models significantly (Table 6). After adding test-level covariates, the marginal R^2 for fluency and originality raises to .14 ($p < .001$) which can be considered moderate. The marginal R^2 for flexibility climbs to .30 ($p < .001$) which is considered substantial. Adding preschooler characteristics and differential intervention effects doesn't improve the models significantly ($p > .05$).

Table 6 Model comparison of the full models

Fluency	*npar*	*AIC*	*-2LL*	*Chi²*	*DF*	*Pr(>Chi²)*	R^2_{cond}	R^2_{marg}
Model 1c: Basic Model	9	652.00	634.00				*1*	.04
Model 2: Test conditions	**13**	**626.39**	**600.39**	**33.61**	**4**	**<.001**	*1*	**.14**
Model 3: Preschooler characteristics	16	628.86	596.86	3.53	3	.317	*1*	.15
Model 4: Differential intervention effects	18	632.49	596.49	0.37	2	.833	*1*	.15
Flexibility								
Model 1b: Basic Model	7	676.40	662.40				.44	.12
Model 2: Test conditions	**11**	**608.85**	**586.85**	**75.55**	**4**	**<.001**	**.60**	**.30**
Model 3: Preschooler characteristics	14	607.33	579.33	7.53	3	.057	.60	.33
Model 4: Differential intervention effects	16	610.18	578.18	1.14	2	.565	.60	.33
Originality								
Model 1c: Basic Model	8	1154.20	1138.20				.67	.05
Model 2: Test conditions	**12**	**1142.30**	**1118.30**	**19.87**	**4**	**<.001**	**.67**	**.14**
Model 3: Preschooler characteristics	15	1144.00	1114.00	4.38	3	.224	.68	.15
Model 4: Differential intervention effects	17	1147.50	1113.50	0.49	2	.782	.68	.16

Note. npar = number of parameters in the model; AIC = Akaike Information Criteria, -2LL = deviance; Chi² = Δ-2LL; DF = degrees of freedom for the likelihood ratio test; Pr(>Chi²) = p-value of the Chi². Bold = selected models. $R^2_{cond\ (fluency)} = 1$ since we fixed the residual variance to near zero in order to allow the model to estimate varying variance on the child level.

Two outliers were removed for fluency and three outliers for flexibility and originality. Concerning fluency and flexibility, inspection of the data revealed that these children scored extreme on the pretest which probably caused a ceiling effect. The outliers in the originality model came from children that were both shy or distracted and not interested, which undermine reliability of the test score. Residuals were sufficiently normally distributed. No multicollinearity and, concerning originality, overdispersion was detected.

Parameter Estimates

In this section, we present the parameter estimates of the final models for the three factors separately. All parameter estimates are presented in Table 7.

Fluency

The treatment had no significant main effect on the fluency score, indicating that both treatment and control group are equivalent at base level (β = 0.01, *SE* = 0.33, *t* = 0.02). Overall pretest and posttest scores did not significantly deviate from zero ($|t| < 1.96$). Recall that fluency scores are standardised, making zero correspond to the mean. Most important, children from the treatment group improved significantly more from the pretest to the posttest than children from the control group (β = 0.48, *SE* = 0.17, *t* = 2.77). In addition, the picture that was used during the test, turned out to have a limited but significant effect on the fluency scores. Use of the test with the ball in the pond resulted in an increase of 0.17 standard deviations (*SE* = 0.08, *t* = 2.11) in comparison to the test with the ball on the cupboard. Finally, shyness (β = -0.84, *SE* = 0.19, *t* = -4.48) and a lack of interest (β = -0.70, *SE* = 0.23, *t* = -3.08) are associated with substantial and significant lower fluency scores.

Flexibility

Similar to fluency, there is no significant main effect of the treatment on the flexibility score (β = -0.03, *SE* = 0.22, *t* = -0.15). Both groups are equivalent at pretest level. Concerning flexibility, pretest scores are significantly lower than the overall average score (β = -0.39, *SE* = 0.19, *t* = -2.04). Posttest scores did not deviate significantly from the mean ($|t| < 1.96$). Again, children from the treatment group improved significantly more than children from the control group (β = 0.40, *SE* = 0.17, *t* = 2.45). As was the case with fluency, the picture that was used for the test affected flexibility scores (β = 0.64, *SE* = 0.08, *t* = 8.04) and children that were shy (β = -0.96, *SE* = 0.18, *t* = -5.42) or had a lack of interest (β = -0.62, *SE* = 0.22, *t* = -2.80) scored below average.

Originality

In contrast to the fluency and flexibility scores, the originality scores are expressed on the original score scale. The simple treatment effect doesn't deviate significantly from zero, indicating equivalence at pretest level between both groups. From the estimate of the intercepts for both the pretest and the posttest we learn that a child from the control group that took the test with the ball on the cupboard and was not shy or distracted, nor showed a lack of interest, is expected to score 2.38 (*SE* = 0.65, *z* = 3.15) for originality on the pretest and 2.15 (*SE* = 0.53, *z* = 3.07) on the posttest. This corresponds with approximately one original solution (given by less than 10 % of the children) or two semi-original solutions (given by 10 to 25 % of the children). The interaction effect between treatment and posttest is significant, indicating that on average, the children in the treatment group improved significantly more on originality than those in the control group. Children from the treatment group scored 1.53 times higher on the posttest than children from the control group (*SE* = 0.32, *z* = 2.04). Shyness (β = -0.50, *SE* = 0.12, *t* = -2.85) and a lack of interest (β = -0.23, *SE* = 0.12, *t* = -2.92) have a negative impact on the originality score. This indicates that children for whom the teacher marked that they were shy, scored approximately 0.50 times lower for originality than children that were not shy. Children who showed a lack of interest, scored 77 % less (0.23 – 1 = -0.77).

Table 7 Parameter estimates for the three final models

	Fluency (N_{Kid} = 130)			**Flexibility** (N_{Kid} = 129)			**Originality** (N_{Kid} = 122)		
Random effects	Var	SD		Var	SD		Var	SD	
Child				0.26	0.51				
Pretest	0.77	0.88		*NA*	*NA*		0.68	0.83	
Posttest	0.54	0.73		*NA*	*NA*		0.14	0.37	
Class	0.15	0.39		0.05	0.23		0.11	0.33	
Residual	*NA*	*NA*		0.34	0.58		*NA*	*NA*	
Fixed effects	Est.	SE	*t-value*	Est.	SE	*t-value*	Est. (exp)	SE	*z-value*
Pretest	-0.09	0.27	-0.34	**-0.39**	0.19	-2.04	**2.38**	0.65	3.15
Posttest	-0.24	0.26	-0.90	-0.26	0.19	-1.35	**2.15**	0.53	3.07
Treatment	0.01	0.33	0.02	-0.03	0.22	-0.15	1.11	0.36	0.31
Picture (pond)	**0.17**	0.08	2.11	**0.64**	0.08	8.04	1.03	0.10	0.27
Shy	**-0.84**	0.19	-4.48	**-0.96**	0.18	-5.42	**0.50**	0.12	-2.85
Distracted	-0.23	0.22	-1.07	-0.14	0.20	-0.70	1.01	0.24	0.05
Lack of interest	**-0.70**	0.23	-3.08	**-0.62**	0.22	-2.80	**0.23**	0.12	-2.92
Treatment*Posttest	**0.48**	0.17	2.77	**0.40**	0.17	2.45	**1.53**	0.32	2.04

Note. Bold typeface indicates $|t| \geq 1.96$ or $|z| \geq 1.96$.

Contrasts between the group means indicate that there is a significant difference between the mean fluency score on the pretest and posttest of the treatment group (*contrast* = - 0.33, *t* = -3.33) (Table 8). The same applies for flexibility (*contrast* = - 0.53, *t* = -5.54) and originality (*contrast* = - 0.33, *z* = -2.61). However, there is no significant difference between the mean of posttest scores from the treatment group and the control group.

Table 8 Contrasts between group means

Fluency	Group	Est.	*SE*	df	*t*-ratio
Pretest – posttest	Treatment group	**-0.33**	0.10	136.00	-3.33
Control – treatment	Posttest	-0.48	0.35	12.40	-1.37
Flexibility					
Pretest – posttest	Treatment group	**-0.53**	0.10	135.00	-5.54
Control – treatment	Posttest	-0.37	0.25	15.20	-1.49
Originality					*z*-ratio
Pretest – posttest	Treatment group	**-0.33**	0.13	Inf	-2.61
Control – treatment	Posttest	-0.53	0.29	Inf	-1.85

Note. Bold typeface indicates $|t| \geq 1.96$ or $|z| \geq 1.96$.

The effect size of the intervention for fluency (*Cohen's d* = 0.57) and flexibility (*Cohen's d* = 0.61) is moderate, according to Cohen (1988).

Discussion

Interactive picturebook reading is an easily accessible activity that provides preschoolers with opportunities to practice problem-solving skills (Gosen et al., 2015). Working on problem-solving skills from an early age, lays a good foundation for future learning (OECD, 2013). By means of a pretest-posttest quasi-experiment with non-equivalent control group, we investigated the effect of interactive picturebook reading on the fluency, flexibility and originality of solutions preschool children can come up with when faced with a visual presented problem.

The results indicate that interactive picturebook reading has a significant and moderate effect on the fluency and flexibility with which preschool children come up with solutions for a visual presented problem. According to Cohen (1988), this means that the effect is large enough to be visible in the course of normal experience. In other words, we expect that preschool teachers will notice an effect of interactive picturebook reading on the extent to which preschoolers are able to think fluently and flexibly when thinking of possible solutions for a problem. Concerning originality, children that received the interactive picturebook intervention also improved significantly more than children from the control group who did not receive the intervention.

Analysis revealed that the picture that was used for the test had a significant effect on the fluency and flexibility score. Children came up with more imaginative solutions in the test with the ball in the pond and also combined more objects or actions. It is possible that the picture with the pond appealed more to the imagination of the children and the picture with the cupboard related more to a real classroom environment. Since most of the picturebooks used during the interventions contained imaginative elements, the intervention may have had more effect if the picture with the pond was used for the posttest.

Another factor that strongly influenced the test results had to do with the reactions of the children to the test taking itself. In agreement with the findings of Rathé et al. (2016), preschoolers whose teacher indicated that they were shy or nervous during the test scored significantly lower on all three factors. The influence of shyness on the test results can possibly be limited by repeatedly administering a test to the children. A lack of interest in the test has also a significantly negative effect on the results. As in the study by Tsamir et al. (2010), some children indicated that they had enough after a few solutions and did not want to search any further. Before the tests, children were sometimes taken out of their play. As a result, they may not have been fully engaged in the test. Being quickly distracted does not seem to have any effect. Perhaps teachers are sufficiently experienced in recovering the children's attention.

Implications, Limitations and Further Research

The current study contributes to the extensive body of scientific evidence on the use of picturebooks to stimulate children's language development (Dowdall et al., 2020) and early mathematical skills (Op ‘t Eynde et al., 2022; Zhang et al., 2023) by demonstrating the effectiveness of using picturebooks to improve preschoolers’ problem-solving skills. Building on the research of Gosen et al. (2015) on the nature of problem-solving interactions during shared reading, we found that discussing possible solutions during picturebook reading is an effective way for young children to practice problem-solving skills. Consistent with Wiseman (2011), this research supports the idea that interactive picturebook reading provides opportunities for children to learn complex skills together in a safe environment.

The findings in this report need to be seen in the light of certain limitations. First, in this study, we only examined the effect of interactive picturebook reading on one of the first steps in the problem-solving process: generating multiple possible solutions to a given problem. Interactive picturebook reading can stimulate flexible and creative thinking, but picturebooks do not offer problems to be solved in the real world. Further research is needed to investigate effective learning activities to enhance other steps of the problem-solving process, such as becoming aware of a problem, exploring and understanding the problem, weighing up various possible solutions against each other, planning and executing a solution, monitoring the process and reflecting on it. In addition, it is still unclear to what extent the skills demonstrated during interactive picturebook reading can be transferred to real-life problem solving.

Second, although we asked teachers about their perception of the problem-solving interactions after every session of the intervention, it was not possible to analyse the effect of the quality of the interactions. Morrow and Smith (1990) pointed out that the nature and amount of interactions during shared reading influences story comprehension. Gosen et al. (2015) found that the selection of the picturebook influences the problem-solving interactions. Future research can explore the conditions for effective interactive picturebook reading to enhance preschoolers’ problem-solving skills.

Third, the entire experiment was carried out by the preschool teachers themselves. In doing so, we tried to enhance ecological validity and to limit the reactions of the children to the presence of an unknown researcher. The downside is that, despite briefings, protocols, instructional videos and training, there may be minor differences in the way the intervention and the test takings were carried out. Due to the COVID-19 pandemic, researchers were not allowed to enter the class for direct observations. We have tried to

compensate for this with research-based logbooks in which teachers could make notes about specific events during the interventions and test-taking.

The results of this study also show that young children are sensitive to unfamiliar situations. Reliable and valid assessment of complex cognitive skills is a major concern in research on the effectiveness of learning activities, especially when young children are involved. To date, no systematic research has been done on possible ways to assess problem-solving skills in the real classroom context. Insights from such a study could contribute to further effectiveness research and provide tools for teachers who want to strengthen their preschoolers' problem-solving skills through assessment for learning (Wiliam & Thompson, 2008).

Conclusion and Practical Recommendations

The current study demonstrates that interactive reading from picturebooks is an effective way to work on problem-solving skills in preschool. Picturebooks are not too expensive and can be found in almost every preschool classroom. They are familiar, entertaining, often beautiful designed and funny. Picturebooks speak the language of children in words and images and can be used in a wide range of approaches that fits the needs and possibilities of preschool children (Fitton et al., 2018). Many picturebooks entail problem situations, making them particularly suitable for practising problem-solving skills.

We conclude with some practical recommendations for applying interactive picturebook reading in the classroom to enhance preschoolers' problem-solving skills. These recommendations are based on the findings discussed in this paper, the findings of previous studies that were used to develop the intervention protocol and the comments and suggestions made by the participating teachers in the logbooks and during the debriefing.

The logbooks show that both preschoolers and teachers need some time to get used to interactive picturebook reading. It takes some practice to elicit interaction with and between preschoolers during shared reading. As the reading sessions progressed, the teachers reported greater satisfaction with the reading sessions and noted that the preschoolers formulated more spontaneous and varied solutions.

During interactive picturebook reading, it is vital that all children have the opportunity to be part of the discussion. Mutual dialogue allows them to build on each other's ideas, generate alternative solutions, elaborate on proposed solutions and evaluate possibilities. Consistent with previous research, we therefore recommend to read interactively from picturebooks in small groups. In addition, we suggest to create heterogeneous groups, mixing children with more and less language proficiency, children with more and less confidence in verbal expression and younger children with older ones. Several participating teachers argue that this encourages less talkative preschoolers to contribute to the discussion. Regularly changing the composition of the groups can also help to maximize each child's opportunity to participate in the discussion.

When choosing picturebooks, it is important to ensure that they are appropriate to the developmental and linguistic level of the children (see also Nicolopoulou et al., 2023) and that the story and visual presentation of the picturebooks appeal to them. Participating teachers noted that not all books are equally suitable for eliciting problem-solving discussion between preschoolers. The story about *The princess with the long hair*, in which a princes has to find a way to carry her immensely long and heavy hair, turned out to be quite difficult for the children. The books of *Cop and Robber* (in which a police officer has to capture a comical fugitive) and *Deep-sea doctor Derek* (in which an inventive doctor has to help impaired and injured aquatic animals) proved to be very suitable.

Declarations

Funding. This research was supported by the Research Foundation – Flanders (Fonds Wetenschappelijk Onderzoek – Vlaanderen) [grant number 1S03523N].

Conflict of interest. The authors report there are no competing interests to declare.

Ethical Approval. The research was approved by the Ethical Advisory Committee for Social and Human Sciences (EASHW) of the University of Antwerp on 12 January 2021 (file SHW_20_112).

Informed consent. Parental consent forms were distributed and obtained along with an information letter and the questionnaire on child characteristics.

Acknowledgement

This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s10643-023-01542-3. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/acceptedmanuscript-terms.

Children's Literature Cited

Picturebooks marked with an asterisk were included in the proposed list.

*Bright, R. & Field, J. (2017). *Twee vechtende eekhoorntjes [The squirrels who squabbled]*. Gottmer.

*Daywalt, D. & Jeffers, O. (2014). *De krijtjes staken [The day the crayons quit]*. De Fontein.

*de Beer, H. (2014). *Haas en Mol zoeken een uitweg [Hare and Mole looking for a way out]*. De Vier Windstreken.

Flas, I. & Masson, A. (2012). *De sleutel [The key]*. Clavis.

*Hudson, K. (2017). *Veel te veel wortels [Too many carrots]*. Ars Scribendi.

*Jeffers, O. (2011). *Vast [Stuck]*. Pimento.

Lloyd, S. & Tickle, J. (2004). *Happende haaien [Yummy yummy! Food for my tummy!]*. Veltman Uitgevers.

*Robinson, M. & Powell, C. (2018). *Hoe verstop je een giraffe? [Have you seen my giraffe?]*. De Fontein.

*Schubert, I., & Schubert, D. (1998). *Een gat in mijn emmer [A hole in my bucket]*. Lemniscaat.

*Stein, M. & van Hout, M. (2006). *Van Mij! [Mine!]*. Lemniscaat.

*Timmers, L. (2015). Garage Gust [Gu's garage]. Querido.

*Timmers, L. (2017). *Diepzeedokter Diederik [Deep-sea doctor Derek]*. Querido.

Van Genechten, G. (2015). *Rikkie heeft een plan [Rikkie has plan]*. Clavis

*Van Haeringen, A. (2015). *De prinses met de lange haren [The princess with the long hair]*. Leopold.

*Veldkamp, T. & de Boer, K. (2008). *Agent en Boef [Cop and Robber]*. Lemniscaat.

Velthuijs, M. (2017). *Kikker vindt een schat [Frog finds a treasure]*. Leopold.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

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

Baars, M., Wijnia, L., & Paas, F. (2017). The association between motivation, affect, and self-regulated learning when solving problems. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01346

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University.

Capin, P., Walker, M. A., Vaughn, S., & Wanzek, J. (2018). Examining how treatment fidelity is supported, measured, and reported in K–3 reading intervention research. *Educational Psychology Review*, *30*(3), 885-919. https://doi.org/10.1007/s10648-017-9429-z

Casey, B., Erkut, S., Ceder, I., & Young, J. M. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology*, *29*(1), 29-48. https://doi.org/10.1016/j.appdev.2007.10.005

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685-709. https://doi.org/10.1007/s11336-013-9328-2

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2 ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8 ed.). Routledge.

Csapo, B., & Funke, J. (Eds.). (2017). *The nature of problem solving: Using research to inspire 21st century learning*. OECD Publishing. https://doi.org/10.1787/9789264273955-en.

Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2020). Shared picture book reading interventions for child language development: A systematic review and meta-analysis. *Child Development*, *91*(2), 383-399. https://doi.org/10.1111/cdev.13225

Fettig, A., Schultz, T. R., & Ostrosky, M. M. (2016). Storybooks and beyond: Teaching problem solving skills in early childhood classrooms. *Young Exceptional Children*, *19*(3), 18-31. https://doi.org/10.1177/1096250615576803

Fitton, L., McIlraith, A. L., & Wood, C. L. (2018). Shared book reading interventions with english learners: A meta-analysis. *Review of Educational Research*, *88*(5), 712-751. https://doi.org/10.3102/0034654318790909

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

Gosen, M. N. (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten* [Doctoral dissertation, Rijksuniversiteit Groningen]. Groningen, Netherlands. https://www.rug.nl/staff/m.n.gosen/binnenwerk_final_mngosen.pdf

Gosen, M. N., Berenst, J., & de Glopper, K. (2015). Problem-solving during shared reading at kindergarten. *Classroom Discourse*, *6*(3), 175-197. https://doi.org/10.1080/19463014.2014.991339

Green, A. J. K., & Gilhooly, K. (2012). Problem solving. In N. Braisby & A. Gellatly (Eds.), *Cognitive Psychology* (2 ed., pp. 301-333). Oxford University Press.

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Hox, J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge. https://doi.org/10.4324/9781315650982

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2 ed.). Sage Publications. https://doi.org/10.4135/9781412984522

Kemple, K. M., Lee, I. R., & Harris, M. (2016). Young children's curiosity about physical differences associated with race: Shared reading to encourage conversation. *Early Childhood Education Journal*, *44*(2), 97-105. https://doi.org/10.1007/s10643-014-0683-0

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Leikin, M. (2012). The effect of bilingualism on creativity: Developmental and educational perspectives. *International Journal of Bilingualism*, *17*(4), 431-447. https://doi.org/10.1177/1367006912438300

Lenth, R. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. (1.7.5) [R package]. https://CRAN.R-project.org/package=emmeans

Lüdecke, D., Ben-Shachar, M., S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Massey, S. L. (2004). Teacher–child conversation in the preschool classroom. *Early Childhood Education Journal*, *31*(4), 227-231. https://doi.org/10.1023/B:ECEJ.0000024113.69141.23

Mayer, R. E. (1994). Problem solving. In M. W. Eysenck (Ed.), *The Blackwell dictionary of cognitive psychology* (pp. 284-288). Blackwell.

Merga, M. K. (2017). Interactive reading opportunities beyond the early years: What educators need to consider. *Australian Journal of Education*, *61*(3), 328-343. https://doi.org/10.1177/0004944117727749

Meyer, L. A., Wardrop, J. L., Stahl, S. A., & Linn, R. L. (1994). Effects of reading storybooks aloud to children. *Journal of Educational Research*, *88*(2), 69-85. https://doi.org/10.1080/00220671.1994.9944821

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979-1007. https://doi.org/10.3102/0034654309332561

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364-386. https://doi.org/10.1177/1094428106291059

Morrow, L. M., & Smith, J. K. (1990). The effects of group size on interactive storybook reading. *Reading Research Quarterly*, *25*(3), 213-231. https://doi.org/10.2307/748003

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nicolopoulou, A., Hale, E., Leech, K., Weinraub, M., & Maurer, G. (2023). Shared picturebook reading in a preschool class: Promoting narrative comprehension through inferential talk and text difficulty. *Early Childhood Education Journal*. https://doi.org/10.1007/s10643-023-01497-5

OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy* [Book Report]. https://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf

Op ‘t Eynde, E., Depaepe, F., Verschaffel, L., & Torbeyns, J. (2022). Shared picture book reading in early mathematics: A systematic literature review. *Journal für Mathematik-Didaktik*. https://doi.org/10.1007/s13138-022-00217-7

Osborn, A. F. (1953). *Applied imagination. Principles and procedures of creative thinking*. Scribners.

Pantaleo, S. (2007). Interthinking: Young children using language to think collectively during interactive read-alouds. *Early Childhood Education Journal*, *34*(6), 439-447. https://doi.org/10.1007/s10643-007-0154-y

R Core Team. (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. (4.2.1) https://www.R-project.org/

Rathé, S., Torbeyns, J., Hannula-Sormunen, M. M., & Verschaffel, L. (2016). Kindergartners’ spontaneous focusing on numerosity in relation to their number-related utterances during numerical picture book reading. *Mathematical Thinking and Learning*, *18*(2), 125-141. https://doi.org/10.1080/10986065.2016.1148531

Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories - Research - Applications*, *4*(2), 238. https://doi.org/10.1515/ctra-2017-0013

Sanetti, L. M. H., Cook, B. G., & Cook, L. (2021). Treatment fidelity: What it is and why it matters. *Learning Disabilities Research & Practice*, *36*(1), 5-11. https://doi.org/10.1111/ldrp.12238

Towson, J. A., Fettig, A., Fleury, V. P., & Abarca, D. L. (2017). Dialogic reading in early childhood settings: A summary of the evidence base. *Topics in Early Childhood Special Education*, *37*(3), 132-146. https://doi.org/10.1177/0271121417724875

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

van den Heuvel-Panhuizen, M., Elia, I., & Robitzsch, A. (2016). Effects of reading picture books on kindergartners' mathematics performance. *Educational Psychology*, *36*(2), 323-346. https://doi.org/10.1080/01443410.2014.963029

van Elsäcker, W., & Verhoeven, L. (1997). Kleuters leren meer van voorlezen in kleine groepen [Preschoolers learn more from reading aloud in small groups]. *Pedagogische Studiën*, *74*, 117-129. http://objects.library.uu.nl/reader/index.php?obj=1874-205429&lan=en#page//99/85/45/998545344215563417009662481179262141 25.jpg/mode/1up

van Kleeck, A. (2003). Research on book sharing: Another critical look. In A. van. Kleeck, S. A. Stahl, & E. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 271–320). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410607355

Van Elsen, J. (2020, december 26). *Take the test: instructiefilm* [Video]. YouTube. https://www.youtube.com/watch?v=0LzRdlpGCTo

Walsh, R. L., & Hodge, K. A. (2018). Are we asking the right questions? An analysis of research on the effect of teachers' questioning on children's language during shared book reading with young children. *Journal of Early Childhood Literacy*, *18*(2), 264-294. https://doi.org/10.1177/1468798416659124

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. (3.3.6) [R package]. Springer-Verlag. https://ggplot2.tidyverse.org

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The Future of Assessment: Shaping Teaching and Learning* (pp. 53–82). Routledge. https://doi.org/10.4324/9781315086545

Wiseman, A. (2011). Interactive read alouds: Teachers and students constructing knowledge and literacy together. *Early Childhood Education Journal*, *38*(6), 431-438. https://doi.org/10.1007/s10643-010-0426-9

Zhang, Q., Sun, J., & Yeung, W.-y. (2023). Effects of using picture books in mathematics teaching and learning: A systematic literature review from 2000–2022. *Review of Education*, *11*(1), e3383. https://doi.org/https://doi.org/10.1002/rev3.3383

Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement*, *19*(2), 149-154. http://www.jstor.org/stable/1434908