Skillen, Johanna; Ricken, Gabi; Seitz-Stein, Katja

Game-based assessment of early mathematical competencies. A pilot study

Empirische Sonderpädagogik 15 (2023) 1, S. 38-60

Empirische Sonderpädagogik, 2023, Nr. 1, S. 38-60
ISSN 1869-4845 (Print) · ISSN 1869-4934 (ebook)

Game-Based Assessment of Early Mathematical Competencies – a Pilot Study

Johanna Skillen[1], Gabi Ricken[2] und Katja Seitz-Stein[1]

[1] Catholic University of Eichstätt-Ingolstadt
[2] University of Hamburg

Abstract
This study examined a game-based screening tool for measuring early mathematical competencies. Since these competencies have been shown to be predictors of later mathematical achievement, assessment tools are required that identify children who are at risk of experiencing developmental difficulties in this area. Game-based approaches have been effective in nurturing early mathematical competencies, but only few studies have examined their potential in assessment. Tasks that address prognostic, relevant competencies in accordance with two developmental models were embedded in a number board game. A sample of 300 German four-to-six-year-olds played the game. We compared the children's performance with that of standardized tests. The results confirmed the applicability of the approach and indicated satisfactory internal consistency and content as well as concurrent validity. Moreover, the screening tool identified about 80 % of the children whose performance was clearly below average. The authors discuss the advantages and disadvantages of game-based assessment.

Keywords: early childhood education, mathematical competencies, screening, number board games.

Pilotierung eines spielbasierten Verfahrens zur Erfassung basisnumerischer Kompetenzen

Zusammenfassung
Die Studie untersucht, ob basisnumerische Kompetenzen mithilfe eines spielbasierten Screeningverfahrens erfasst werden können. Da sich diese Kompetenzen als bedeutsam für schulische Mathematikleistungen erweisen, besteht ein Bedarf an Verfahren, mit denen Entwicklungsschwierigkeiten frühzeitig erkannt werden können. Spielbasierte Verfahren haben sich in der Förderung als wirksam erwiesen. Für die Diagnostik wurde ihr Potential noch wenig überprüft. Ausgehend von zwei Entwicklungsmodellen wurden Aufgaben ausgewählt, die prognostisch relevante mathematische Kompetenzen adressieren und in ein Zahlenbrettspiel eingebettet. Eine Stichprobe von 300 deutschen Kindergartenkindern im Alter von 4 bis 6 Jahren spielte das Spiel. Zum Vergleich wurden standardisierte Testverfahren eingesetzt. Die Ergebnisse bestätigen die Durchführbarkeit und sprechen für eine zufriedenstellende Reliabilität, sowie für Inhalts- und konkurrente Validität. Mithilfe des Screenings konnten rund 80 % der Kinder als leistungsschwach identifiziert werden, die in

den standardisierten Verfahren altersabweichende Leistungen zeigten. Stärken und Schwächen einer spielbasierten Erfassung werden diskutiert.

Schlüsselwörter: frühkindliche Bildung, basisnumerische Kompetenzen, Screening, Zahlenbrettspiele

In recent decades, the description and support of young children's development in numerical competence has become a major part of early childhood research and education with a particular focus on preparation for school and later academic success. Hence, there is a need for concepts and materials that allow practitioners to fulfil their educational mandate (e.g. Bock-Famulla et al., 2015; Textor, 2005). In order to meet the demands of early childhood education, resource-saving and low-threshold diagnostic and training approaches are needed (Jörns et al., 2013; Seeger et al., 2014). Recently, game-based approaches have gained increasing research interest as tools in cognitive training and scholastic learning, especially in mathematics education (see Bayeck, 2020; Boyle et al., 2016 for systematic reviews). Games are considered an attractive and less resource-intensive alternative to traditional instructional approaches (Hirsh-Pasek et al., 2009; Weisberg et al., 2016). In contrast to free play, they involve previously codified rules, which can be communicated explicitly. Many games require competitive activity and can be challenging (Baines & Blatchford, 2011). In addition, positive characteristics of play can also be found, for instance, fun, interactivity, engagement, or opportunities to demonstrate the players' abilities in terms of self-efficacy (Hassinger-Das et al., 2017). Several studies have shown that the gamification of cognitive assessment and training tasks can increase motivation, engagement and performance across various age groups (for a review see Lumsden et al., 2016; Ninaus et al., 2015). Hence, games provide a setting that allows for pleasant but also guided and therefore focused (learning) activities (Hauser et al., 2014). So far, different studies have investigated the conception and conditions of rule-based games, in order to promote improvements in early mathematical competencies at preschool age (e.g. Gasteiger, 2013; Häsel-Weide et al., 2014; Jörns et al., 2014; Ramani & Siegler, 2008; Wittmann & Müller, 1995).

A question that has only been addressed by a few studies so far is whether the effective and potentially low-threshold game-based approaches may not only be useful in promoting mathematical competencies, but also in assessing them. Games may help to access mathematical competencies in young children and children with special educational needs, in a way that more artificial, traditional assessment settings might not. We see motivational, emotional and social advantages in game-based assessment. In particular, children who normally refrain from engaging in mathematical activities might be willing to take part in a game. The setting resembles a natural play situation rather than a traditional assessment, offering children the chance to win without reference to their actual mathematical performance. Hence, the game-based assessment provides fun and interaction for all children 'on a level playing field', regardless of their mathematical abilities or whether they are playing against another child or an adult. This could reduce negative emotional states during the assessment itself. Hence, engagement in these activities might be facilitated through use of the game-based assessment tool. However, children have to process game requirements while carrying out mathematical activities, which strains cognitive resources and could reduce mathematical performance. In order to investigate whether game-based assessment of mathematical competencies can be applied to young children, we developed a game-based screening tool. In the current pilot study, we report on the conception of our number board game *House of Numbers* (HoN) and on the first validation of this game as innovative assessment tool.

Development of mathematical competencies

The acquisition of mathematical competencies begins well before formal education. Several authors consider the ability to discriminate numerical magnitudes in infancy as core systems of mathematical competence (Aster & Shalev, 2007; Feigenson et al., 2004). This includes, for example, the non-symbolic approximate comparison of two dot arrays (Aunio & Räsänen, 2016). Fischer and colleagues (2017) differentiate seven further basic competencies in early mathematical development. Children learn how to count by acquiring the number word sequence and mapping number words to quantity elements via 1-to-1 counting principles. This implies an ordinal concept of numbers as sequenced elements with fixed positions. Moreover, children begin to process number symbols and learn how to read and write numbers. A central milestone is the acquisition of the cardinal value of number words. Children can understand that numbers represent specific magnitudes. This precise representation of numbers allows for symbolic magnitude comparisons meaning that children can compare two digits according to their size (Aunio & Räsänen, 2016; Schneider et al., 2017). Subsequently, they can acquire more sophisticated competencies and understand relations between numbers. Children can understand the decimal system, operate in different number ranges (Fischer et al., 2017), and solve arithmetical problems (Chu et al., 2018).

In view of the precise developmental process of early mathematical competencies, several models can be considered in the German-speaking context (e.g. Aster & Shalev, 2007; Ricken et al., 2013; Krajewski, 2008). They mostly concur on the mathematical concepts children acquire, but differ in hierarchical composition (Fischer et al., 2017). For example, Krajewski (2008) describes the development of mathematical competencies from early childhood to primary school age via three developmental levels (model one; M1 in the following). She assumes that children can initially discriminate quantities, acquire number words as well as the exact number sequence, and get to know numerals. However, they do not yet perceive a relationship or form a link between quantities and numbers. In the second level, children become aware of the relationship between quantities and numbers, understand the meaning of number words, and can allocate numbers to the quantity they represent. The development of this precise quantity-number representation is considered to be the central developmental step. Furthermore, children understand the dimensions of quantities and numbers in relation to each other e.g., 5 is 2 larger than 3. They can compose, decompose, and determine differences between quantities and numbers (level 3). Longitudinal studies provided empirical evidence in support of these three identified levels. Results show that the basic numerical competencies acquired in level 1 account for the majority of variances in level 2 and 3 competencies, while level 2 and 3 competencies are a reliable predictor of later mathematical school achievement (Krajewski & Schneider, 2006; 2009).

A second model (hereafter 'M2') identifies five developmental levels (Ricken et al., 2013). In each of these, children acquire several mathematical abilities based on the understanding of a specific arithmetic concept. The first level comprises counting abilities. Children acquire the ability to follow an exact number sequence and can map numbers to objects using 1-to-1 strategies. In the second level, children acquire a mental number line and are able to name the numbers that precede and follow a given number. At level three, they understand order-irrelevant principles in counting, start to map number line positions to quantities, develop a cardinal magnitude representation, and understand decomposability. The authors consider this third level as central developmental milestone for the acquisition of effective arithmetic strategies, as it enables children to understand further relations between numbers and quantities, such as part-part-whole concepts (level 4) or relationality (level 5). More recently, Fritz and colleagues (2018) added a sixth level that describes the equidistant bundling of numbers as a precondition of multiplication and division. By means of Rasch model analyses, the authors provided empirical evidence for their theoretical assumptions. Within the Rasch model, items are scaled with respect to item-difficulty and person-ability. The model proved that the items group according to the five developmental levels. That suggests that children who can master items at higher levels can also solve lower-level items (Ricken et al., 2013; van Vuuren et al., 2021). Moreover, results of longitudinal studies show that the mathematical development of kindergarten and primary school children follows the five levels, with the individual level of development an effective predictor of mathematical achievement at the end of grade two (Fritz et al., 2018).

Prognostic relevance of early mathematical competencies

Irrespective of the theoretical model and conceptualisation of developmental stages, many studies have shown that early mathematical competencies are a specific and powerful predictor of later, scholastic mathematical achievement (Aunio & Räsänen, 2016; Garon-Carrier et al., 2018; Geary, 2011; Krajewski & Schneider, 2009; LeFevre et al., 2010; Sasanguie et al., 2012). Preschool number and quantity knowledge accounted for about 26 % of the variance in mathematical performance at the end of grades 1 and 4 (Krajewski & Schneider, 2006). Some competencies that are acquired at an early stage in children's mathematical development prove to be of particular importance. Counting skills such as counting forwards and backwards, counting from any given number in the number word sequence and counting off elements of a given quantity highly correlate with later mathematical performance in school, and are crucial for determining differences in performance (Martin et al., 2014; Nguyen et al., 2016). Moreover, several studies show that the cardinality concept and symbolic magnitude representation prove to be prognostically relevant (Cueli et al., 2019; Ferreira et al., 2012; Merkley & Ansari, 2016; Schneider et al., 2017). Non-symbolic approximate magnitude representation, in addition to other domain general predictors such as working memory, seem to influence academic success indirectly. They contribute to the development of children's early mathematical competencies such as cardinal number knowledge (Aragón et al., 2020; Chu et al., 2015; Costa et al., 2018).

There are considerable differences in the early mathematical competencies of children even from very early stages of development (Aunola et al., 2004). The integration of number and quantity concepts seems to be a critical developmental step, which presents some children with severe difficulties (Aunio & Räsänen, 2016; Fritz & Ricken, 2008; Gerster, 2003,). Differences in performance have been shown to remain constant throughout preschool years and persist even after the child enters formal education (Navarro et al., 2012). Therefore, deficits in early mathematical competencies can be associated with later learning impairments (Weißhaupt et al., 2006).

Early detection of developmental difficulties

Considering the prognostic relevance of early mathematical competencies, it seems expedient to detect any deficiencies in young children at an early stage and to provide support for children exhibiting lower performance. Several standardized tests are already available for the assessment of early mathematical competencies. One example is the *Test mathematischer Basiskompetenzen im Kindergartenalter* (MBK 0; Test of Basic Mathematical Competencies in Preschool Children; Krajewski, 2018). It offers a differentiated assessment of children's mathematical abilities in accordance with M1. Krajewski (2018) used classical test theory measures for test evaluation. The MBK 0 provides tasks with a broad range of satisfying item-total correlation and a substantial internal consistency of $\alpha = .96$. The correlation of .67 with another test of early mathematical competencies supports the validity of this instrument. A second example is *Mathematik- und Rechenkonzepte im Vorschulalter - Diagnose* (MARKO-D; Assessment of Math Concepts in Pre-school Aged Children; Ricken et al., 2013), which is based on M2. The test operationalizes the original five developmental levels of the model. It indicates what levels children have already mastered, and which level they are currently acquiring. As stated above, the validation of this test and the underlying developmental model are based on item response theory, with Rasch model analysis as a measure of homogeneity. The model was proved to be valid, with satisfying standardized MNSQ Infit values for all items and a high test or person reliability of .91. A high correlation of .77 with another mathematical test also demonstrates the validity of MARKO-D (Ricken et al., 2013). In addition, translated and adapted versions of the test, as well as the underlying developmental model, have been proved to be valid in different languages and cultures, for example in South Africa (Henning et al., 2019).

Besides these comprehensive tests, screening tools enable the efficient selection of at-risk-children who should receive both a differentiated diagnosis, and training (Deimann & Kastner-Koller, 2021; Tröster, 2009). In correspondence with both of the developmental models described, screening versions of MBK 0 and MARKO-D were developed and tested, and shown to be suitable (Ehlert et al., 2020; Krajewski, 2014).

In general, the well-established tests described here provide child-friendly and to some extent playful tasks and materials, but they do not make use of gamification or actual games. This option was considered by Ninaus and colleagues (2017), who measured the conceptual knowledge of fractions among fifth graders, using the game-based rational number research engine *Semideus*. Students played the game on tablet computers by navigating an avatar along a number line. The assessment was directly implemented into the game. For example, players had to arrange fractions according to their magnitude on the number line. The authors demonstrated that their digital game allows a valid assessment of students' fraction knowledge (Ninaus et al., 2017) and that the game can also be used to foster children's conceptual knowledge of rational numbers (Kiili et al., 2018). Other researchers followed a similar approach and established an online game-based assessment of third and fourth graders' performance in addition and subtraction (Shih et al., 2019) or investigated the validity of an assessment of physics knowledge as well as the effectiveness of in-game supports on learning physics in an educational game (Shute et al., 2021). Moreover, Ponticorvo and colleagues (2021) introduce a physical and digital version of a card game that can be used to assess and train numerical abilities across the lifespan but the game has not yet been tested. Hence, the integration of assessment and training within one game gained increasing research interest but has been focused on school aged children so far. To the best of our knowledge, no game offers comparable advantages in the assessment and education of basic mathematical competencies at kindergarten age.

Conception of the game-based screening tool

The current study is part of an ongoing research project. We are developing a simple game that can be used both as a screening tool and a training tool for early mathematical competencies. In the investigation of games in early childhood education, different research groups make use of analogue games (Ehm et al., 2017), such as conventional board or card games (e.g. Gasteiger & Moeller, 2021; Scalise et al., 2020). These materials can be easily integrated into activities at home and at kindergarten (Lange et al., 2021). Number and magnitude related games have already been developed in accordance with M1 and have been successfully implemented in nursery schools (e.g. Hauser et al., 2014; Jörns et al., 2017). In addition, a developmentally adaptive version of a memory game proved effective in promoting the competencies defined in M2 (Herzog & Fritz, 2022). Both of these game-based approaches comprised several weeks of training for the children and the nursery school teachers. In contrast, linear number board games have been assessed internationally and proved to be effective in different settings and with remarkably few training sessions (Elofsson et al., 2016; Laski & Siegler, 2014; Ramani et al., 2012; Siegler & Ramani, 2009; Whyte & Bull, 2008). We therefore decided to apply this promising approach. Taking into account well-established findings of early number and magnitude competencies in young children, as well as the special characteristics of these games, we developed the number board game 'HoN'.

One main characteristic of linear number board games is the arrangement of numbers in a setup that is closely aligned with the mental representation of numerical magnitudes. In board games containing number ranges above ten, the numbers are arranged in ascending rows. Progressing from left to right, each row consists of ten numbers. Hence, the board embodies the base-ten system with the rows representing decades and the columns representing units (Laski & Siegler, 2014). The design of our number board game uses the well-known image of an ordinary house with ascending floors, each containing ten doors (see Figure 1). The idea is that one number lives behind each door, arranged in a linear setup according to the magnitude. The colour of the floor deepens every two rows to provide an additional cue to the ascending magnitudes. Thus, the HoN reflects the mental representation of numerical magnitudes within a suitable narrative. Further details of the game can be found in the *Method* section of this article, as well as in several training studies concerned with the validation of different training versions of the HoN. The game has already been effective in promoting the early mathematical abilities identified in M1 (Skillen, Berner, Ricken & Seitz-Stein, 2018; Skillen, Berner & Seitz-Stein, 2018). Additionally, a modified version has proved effective in supporting the development of early mathematical competencies defined in M2 (Berner et al., 2021).

The current study focuses on the conception and first validation of the screening version. The aim was to embed a screening tool within the game, while retaining the actual game board and typical in-game actions. It should be possible to objectively, reliably and validly assess whether or not children have already mastered prognostically relevant early mathematical competencies as they play the HoN board game. To fulfil this psychometric requirement, we intended to integrate items that enable a structured, standardized and explicit rating during the course of the game. As both of the previously outlined developmental models are supported by empirical evidence and have been proved suitable, not only in the conception and investigation of different game-based training approaches, but also in the conception of several assessments, we took both of the models into account when constructing the items. We therefore selected approved task types of the standardized tests MBK 0 and MARKO-D (which operationalize the two models) and embedded them in the game with relevant changes to material and setting. Hence, the board game provides the framework or environment, while the actual assessment is based on approved tasks. In the first step, we constructed a broad variety of tasks in line with the two models. This should enable a selection of tasks for the final version of the screening tool on the one hand, and a double validation of the game on the other.

Aims of the present study

The present study pursues three goals. The first aim is to explore the applicability of game-based assessment in young children and to demonstrate the advantages and disadvantages of using a board game as an assessment tool. The question is whether tasks can be administered in a game for children as young as 4 years of age. We investigated the feasibility of the items and item statistics. In order to examine whether processing mathematical tasks while playing a game affects children's performance, we compared achievements in game-based and standardized assessment settings.

The second aim was to evaluate whether the game-based approach enables reliable and valid assessment of early mathematical competencies. Since the item development was theory-based and aligned with approved task types, we expected a consistent item structure, representation of age-related performance differences, and positive correlations with standardized measurements.

Thirdly, we investigated whether low-achieving children can be reliably identified. Since we integrated items addressing prognostically relevant mathematical competencies, we expected a successful detection of low-achieving children.

Method

Participants

The present study included N = 300 children (156 male and 144 female; see Table 1). Recruitment involved 15 German kindergartens (14 in Bavaria and one in Hamburg). The two participation criteria were age-appropriate development based on information provided by educators and sufficient knowledge of the German language.

Informed parental consent was obtained for all participating children. Furthermore, 236 parents informed us about the child's family background via a short questionnaire. Of those, 26 % indicated that at least one of the parents was not born in Germany. At home, 17% of families speak German and at least one other language. With regard to their highest school-leaver's qualification, 43 % of mothers and 39 % of fathers have reached the highest school-leaving qualification in Germany (Abitur).

A group of low-achieving children was formed and consisted of children with

Table 1
Sample statistics

			Age (years; months)			
	n	*m/f*	*Min*	*Max*	*M*	*SD*
4 years	88	50/38	4;1	4;11	4;7	0;3
5 years	101	47/54	5;0	5;11	5;6	0;3
6 years	111	59/52	6;0	6;11	6;3	0;2
entire sample	300	156/144	4;1	6;11	5;6	0;9

scores below a standard *T* score of 40. As two standardized measures of mathematical competencies were administered in this study, two low-achieving groups will be examined. One was detected by the test MBK 0 and comprises 32 children (15 male, *M* = 5 years, 9 months). The other was detected via MARKO-D and includes 19 children (10 male, *M* = 5 years, 5 months).

Research design

We chose a standardized and controlled research design for this pilot study. All children participated in the screening version of the HoN. Furthermore, 275 children took part in the MBK 0 and 276 children completed the MARKO-D. Thus, the majority of children (namely 251) participated in both standardized measurements. The presentation order of assessment tools was balanced between participants. Sixteen trained and supervised research assistants supported data collection as part of their Bachelor thesis in Psychology. All of these research assistants followed the same standardized instructions when conducting the measurements. They received the same introduction to the research project and practiced the use of materials until a standardized form of administration could be assured. For the assessment sessions, children met individually with an experimenter in a separate room in the kindergarten. We presented all measures in separate one-to-one sessions in accordance with the test manuals of the standardized measures. The child sat in front of the material (e.g., the board game) and opposite the experimenter. On average, it took players *M* = 27 minutes (*SD* = 6.7) to complete the screening. MBK 0 lasted *M* = 25 minutes (*SD* = 6.2) and MARKO-D *M* = 28 minutes (*SD* = 6.6).

Material and procedure of the game-based assessment

The screening version of the HoN visualizes a house with five floors; ten doors each (see Figure 1). The resulting 50 doors contain the Arabic numbers 1 to 50 progressing from left to right in each row. The numbers were written using the typical German notation taught in primary school. The entrance is marked with 0 and the 50 with a finish button. When introducing the game to the children, the experimenter explained the number arrangement and told the child that they were going to 'visit' all of the numbers between 1 and 50 living in the HoN. They showed how players move along the ascending floors. Starting at the ground floor, players rolled the dice in turn and moved from door to door by counting their steps from one onwards. The player who reached 50 first won the game. Instead of a conventional dice, a 10-sided dice was used with the Arabic numbers 1 to 5 on twice. On the first three floors (numbers 1 to 30), players use one dice. On the remaining two floors (numbers 31 to 50), they are given a second dice. The use of two dice allows the integration of test items with numbers between one and ten, while moving along floors with higher number ranges.

While progressing through the house, the experimenter verbally provided tasks to the child and recorded the answers. The first version of the screening tool assessed children's mathematical competencies using 38 items. Examples of items and the developmental models they correspond to are provided in Table 2. Tasks were adapted from the standardized measures and embedded in the game. We selected items that address all of the levels of the underlying models. Each competence stated in the models is represented through at least one item. However, we emphasized prognostically relevant or critical competencies in the task selection (i.e. levels 1 and 2 in M1 and levels 1, 2 and 3 in M2). Of the 38 items, 16 items align with competencies defined within both models, while 10 of the items correspond to competencies defined exclusively within M1 and 12 to competencies identified solely in M2. We implemented the chosen tasks as far as possible with

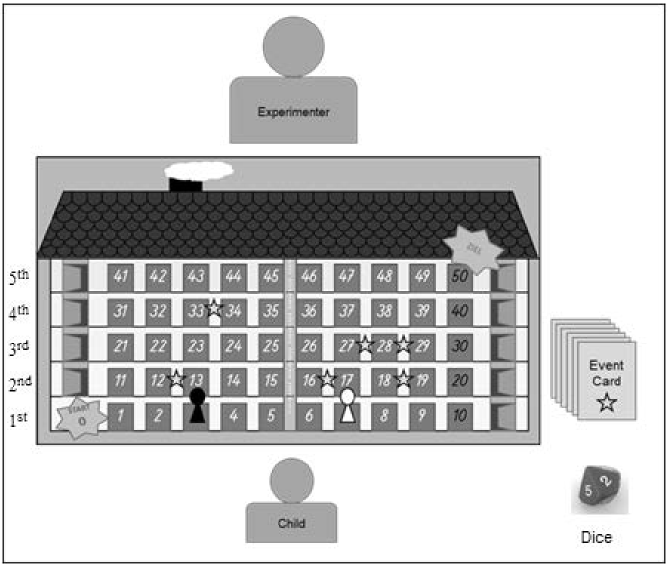


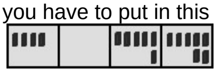
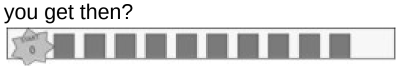
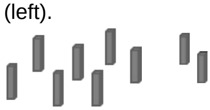
Figure 1
Material and arrangement of the game-based assessment. The game board is displayed in black and white but originally coloured in green, yellow and red.

the material provided by the number board game. However, it was necessary to use additional material in order to realize some of the task types. To preserve the character of the game, we included a typical feature of conventional board games and implemented event cards. Small stars between some of the doors trigger the use of these cards. As soon as one player crossed a star, the child drew an event card and was given a special task. After finishing the task, the child moved forward one step, regardless of the outcome. Generally, children did not receive feedback on any of the mathematical activities (including e.g. to count the steps while moving on) during the game, in order to avoid interference with the assessment.

Standardized instructions indicated the item order and wording to the experimenter. All children received the same 38 prompted screening items in the same order; there was no age-dependent selection of tasks. Moreover, we mixed items of various difficulty levels. Hence, younger children or children of different developmental levels could play the entire game without feeling overstretched. One special characteristic of the game-based setting is that the actual numerical values of 14 tasks depend on the progression of the game. Although each child was given the same test items, the numbers the items related to depended on the position of the child's token or the numbers that the dice showed, which in turn depended on the course of the game and was basically a matter of chance. For example, one item consisted of naming the preceding and the consecutive numbers of a given number. This task was administered while players moved along the first floor. If the token was stood on number 5, the experimenter asked the child to name the preceding and the consecutive numbers relating to 5 (while covering the respective parts of the board). If the token was stood on number 4 or 6, the task was to name the neighbours of 4 or 6. Hence, the children were given the same test item, but in relation to different numbers. To ensure items were comparable in terms of requirements and difficulty, the experimenter was given instructions regarding the number range within which each item had to be administered.

Table 2

Examples of items in the game-based screening, the targeted math competencies, and the underlying models they correspond to

Model	Level	Number of Items	Math Competencies	Item Examples
Model 1 provided by Krajewski (2008)	**Level 1**	11	Counting	[First item. Provided during game introduction.] All the numbers from 0 up to 50 live in the House of Numbers. Do you already know numbers and are you able to count? Please count as far as you can.
			Number identification	[Provided within different number ranges: 1 – 10; 11 – 15; 26 - 29] Do you know the number that the dice shows / your token stands on? Please name it.
	Level 2	9	Precise quantity-number representation	[Provided via an event card while moving along the 2nd floor.] Look, these sticks have a special order. This box is empty. How many sticks do you have to put in this box? 
				[Provided while moving along the 2nd floor.] The dice shows …. Please give me … sticks.
	Level 3	6	First arithmetic	[Provided while moving along the 4th or 5th floor. Children get 10 sticks and have to visualize how they solve the task.] The dice show … and …. How many steps do you have to move on?
Model 2 provided by Ricken et al (2013)	**Level 1**	9	Counting	[Provided during game preparation. Experimenter presents 5 player tokens.] How many player tokens do you see? Please choose 2 to play with.
				[Provided via an event card while moving along the 4th floor.] Please divide the sticks between the two player tokens so that each token gets the same number of sticks.
	Level 2	5	Mental number line	[Provided while moving along the 1st floor. Experimenter covers the numbers around her token.] My token stands on … Which number follows … ?
	Level 3	7	Cardinality and Decomposability	[Provided during game board introduction. Ground floor is covered with the stripe presented below.] (How many numbers live on the first floor?) Now I have an idea. What if you start to count over here? How many doors do you get then? 
				[Provided via an event card while moving along the 2nd floor.] Look at these sticks. Please put here (right) as many sticks as you can see here (left). 
	Level 4	3	Part-Whole Concept	[Provided while moving along the 4th or 5th floor.] You may move on … steps. Please give me … sticks. Of these 3 must be black.
	Level 5	4	Relationality	[Provided while moving along the 1st floor. Experimenter covers the numbers around the child's token.] Your token stands on … Which number is three higher than …?

As the experimenter had already provided the first items during game instruction and preparation (e.g. count the amount of player tokens), the children were involved in the assessment right from the start. Moreover, children always made the first move when starting the actual game. Combined with the rule that children move on one step after playing an event card, it was possible for them to overtake the experimenter or to easily catch up if they fell behind. Hence, the probability of winning the game was biased in favor of the child.

Standardized measures

In order to validate the game-based assessment, we compared the children's performance with that of two standardized measurements, namely MBK 0 and MARKO-D. The experimenter provided the items verbally to the child (in part supported by test material, e.g., images or chips) and recorded the answers in accordance with the test manuals.

MBK 0 includes 58 items and assesses children's basic numerical competencies (level 1), for example in counting forwards and backwards or with numeral identification tasks. Among others, it measures quantity-number representation (level 2) with seriation tasks (requires the child to fill a gap in ascending quantities with the appropriate number of chips). Arithmetic abilities (level 3) are assessed with tasks such as summing up two amounts of chips. In the present study, the internal consistency of all items of the MBK 0 proved substantial with Cronbach's α = 94.

MARKO-D comprises 55 items embedded in a cover story of two squirrels. For instance, the child has to count a number of nuts (level 1) or help the animals in determining the preceding or consecutive number of a given number (level 2). Again, a seriation task operationalizes cardinal magnitude representation (level 3). Another task requires the child to add the missing number of nuts to a total quantity. Moreover, the test provides items requesting the child to compose or decompose quantities (level 4) or to determine the difference between them (level 5). In the present study, the test showed a high person reliability of .92 and an item reliability of .99.

Data analysis

As we implemented items based on two different developmental models, we decided to evaluate them in terms of theory, that is to say separately. Consequently, two different overall scores (M1 score and M2 score; calculated by summing the points achieved) will be considered in the following data analysis.

We examined missing data as an indicator of item administration problems, conducted item analysis via classical test theory (item difficulty and item-total correlation) and item response theory measures (Rasch model), and compared the children's performance by means of correctly solved items in game-based and standardized measures to determine applicability. Cronbach's α as well as person and item reliability via the Rasch model were calculated to test internal consistency. We examined age effects in the children's mathematical competencies and correlations between performance in the screening tool and standardized measurements to investigate validity. Finally, we determined quality indicators for screenings such as specificity, sensitivity, and relative improvement over chance (RIOC-Index). In data analysis we used IBM SPSS Statistics 26, Winsteps 4.3.4, the *Calculation of Test Quality Criteria for Screenings* by Lenhard und Lenhard (2014) and the *Computation of Effect Sizes* by Lenhard und Lenhard (2016).

Results

Applicability

In order to evaluate the applicability of the game-based assessment, we examined the number of missing values in the screening. They occurred due to the children's performance, motivational difficulties, or due to handling problems on the part of the experimenter. Of the 300 children who completed the HoN, 1.7 % revealed performance-related or motivational problems. In this case, the experimenter skipped one or more of the screening items. Handling problems occurred especially in relation to three specific items and gave rise to missing data in the assessments of 10 % of the children. One of these items related to numeral identification in the number range 26 to 29; another task required the child to sum up the numbers shown on the two dice and to visualize the relevant mathematical operations with chips; the third difficult item required the child to name a number that was one more than the number the token was standing on. The experimenter noted that they experienced difficulties conducting the items in the course of the game (e.g. they missed administration or the game was over before the item could be administered). This indicates structural problems as well as a high item density. Nonetheless, we found the screening could be administered to all of the participating children. Hence, the game-based assessment proved to be generally feasible.

In the next step, we conducted item analysis. Similar to MBK 0 and MARKO-D test evaluations, we used classical test theory (M1) and item response theory measures (M2). The majority of the 26 screening items measuring mathematical competencies according to M1 displayed item difficulties between .11 and .97 and an item-total correlation between .20 and .66. One item did not show satisfying values (item difficulty = 1; item-total correlation r_{it} = -.02) and was excluded from data analysis. The 28 items measuring mathematical competencies according to M2 entered a Rasch model, which converged in eight iterations. All items displayed a satisfying MNSQ Infit of 1 ± 0.3 (Wright & Linacre, 1994). The Person/Item MAP showed that the majority of items matched the five developmental levels that were expected based on the theoretical model, with the exception of five items. With regard to these five items, two were found to be more difficult than expected and three were found to be easier. These five items were consequently excluded.

Furthermore, we compared children's performance in game-based and standardized measures to investigate whether the game-based setting affects children's achievement. Mathematical performance as a function of age group and for the entire sample is shown in Figure 2. Evidently, there seemed to be performance differences between age groups but not between assessment settings. To test this, we conducted repeated-measures analyses of variance (ANOVA) with the test instrument as the within-factor and the age group as the between-factor.

With regard to the children's performance in HoN and MBK 0, the ANOVA showed significant main effects for test $F(1, 272) = 7.77$, $p = .006$, $\eta_p^2 = .03$ and age group $F(2, 272) = 53.81$, $p < .001$, $\eta_p^2 = .28$. The interaction was significant, $F(2, 272) = 7.94$, $p < .001$, $\eta_p^2 = .06$. Consequently, the performance of different age groups varied statistically with administration of game-based or standardized measurement. Post-hoc tests revealed fewer correctly solved items in the HoN compared to the MBK 0 for 5-year-olds, $t(100) = -2.82$, $p = .006$, $d = -0.12$ and 6-year-olds, $t(85) = -3.60$, $p = .001$, $d = -0.25$ but no differences in 4-year-olds, $t(87) = 1.46$, $p = .149$. Cohen's d indicated small effects.

Comparison of the children's performance in HoN and MARKO-D indicated significant main effects for test $F(1, 273) = 5.42$, $p = .021$, $\eta_p^2 = .02$ and age group $F(2, 273) = 53.98$, $p < .001$, $\eta_p^2 = .28$ but no

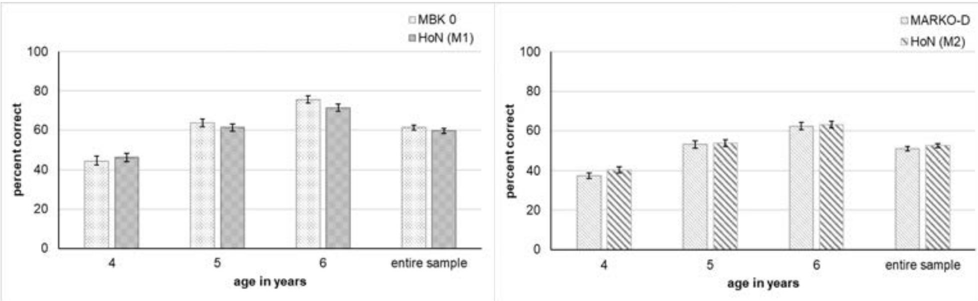


Figure 2
Mathematical performance in MBK 0 vs. HoN (M1) (left graph) and in MARKO-D vs. HoN (M2) (right graph) as a function of age group and for the entire sample. The graphs show the mean percent correct and standard errors.

significant interaction, $F(2, 273) = 1.21$, $p = .301$. Hence, the children correctly solved slightly more items in the game-based assessment than in the standardized measurement across all age groups. Again, the effect size of $\eta_p^2 = .02$ is small.

To conclude, the investigation of missing data indicated difficulties in carrying out the game-based assessment in a small percentage of cases. Item analysis revealed that the item characteristics were suitable for the majority of the tasks implemented. Performance between game-based and standardized assessments showed a statistically significant difference. However, the effectual difference was minimal in terms of practical significance.

Reliability and validity

To evaluate reliability, we examined the internal consistency of the embedded screening items. With regard to the remaining 25 M1 items, Cronbach's α was substantial with a score of .89. With regard to the remaining 23 M2 items, the Rasch model stated a substantial person reliability of .81 and item reliability of .99.

We assessed validity by investigating whether mathematical performance differed between age groups (i.e. 4, 5 and 6-year-old children) in the game-based approach. Since the two underlying developmental models posit considerable changes in mathematical performance during early childhood, a valid assessment tool should reflect age effects. Indeed, mean performance was lowest for younger children and highest for older ones (see Figure 2). Univariate ANOVA with age group as the fixed factor and mathematical performance as the dependent variable satisfied the criteria for homogeneity of variance in the Levene-Test ($p = .390$) and revealed a significant age effect for the M1 score, $F(2, 272) = 42.41$, $p < .001$, $\eta_p^2 = .24$. In contrast, the M2 score did not achieve the criteria for homogeneity of variance ($p = .030$) and Welch's F test was used. The ANOVA indicated a significant age effect as well, Welch's $F(2, 181.55) = 49.94$, $p < .001$, $\eta_p^2 = .98$. Post-hoc comparisons using either the Bonferroni or the Games-Howell procedure confirmed significant differences in early mathematical competencies between all three age groups for M1 and M2 ($p \leq .001$).

In a second step, we investigated the concurrent relationship between children's mathematical performance in the HoN and the standardized measurements via Pearson product-moment correlations. Results displayed a strong positive correlation between children's performance in the HoN (M1 score) and the MBK 0, $r = .90$, $p < .001$, $n = 275$. The same applies to the correlation between the M2 score and MARKO-D with $r = .84$, $p < .001$, $n = 276$.

Identification of low-achieving children

In respect of the two standardized measurements, 32 children performed below average (scores below average standard *T* score of 40) in the MBK 0; and 19 in the MARKO-D. To investigate whether the game-based assessment correctly identifies these low-achieving children, we determined databased cut-off values via *receiver operating characteristic* (ROC) analyses. For each possible cut-off value, the ROC analysis plots sensitivity (number of children correctly identified as low achieving by the screening) vs. 1-specificity (number of falsely identified children). The resulting ROC curves are displayed in Figure 3. The diagonal indicates the curve in the event that the screening was unable to differentiate between average and below average performing children. The jagged curves represent the actual ROC curves for our newly developed game-based assessment. The area between the actual ROC curve and the diagonal is called *area under the curve* (AUC) and is larger, the better a measurement differentiates.

Since we found age-dependent performance differences in the current sample, the two possible screening scores were *z*-transformed within each age group to arrive at age-adjusted test scores. These scores were entered into two separate ROC analyses as test variables. The classifications of standard *T* scores < or ≥ 40 served as trait variables. With regard to the screening assessment of mathematical performance based on M1, a high AUC value of .94, $p < .001$, 95 % CI [.91, .97] indicated that the game-based assessment has a substantial differentiation capability. Two possible *z*-scores of -0.78 (percentile rank = 22 %) and -0.59 (percentile rank = 28 %) showed a large distance between the ROC curve and the diagonal, demonstrating a satisfying proportion of sensitivity and specificity (number of children correctly identified as achieving average performance by the screening). Using this scores as cut-off values for low achievement in the screening tool resulted in a level of correct identification (sensitivity) of either 81 % or 94 % of those children who were identified as under-achievers in the MBK 0. We achieved a specificity of 84 % vs. 78 %, but also a comparably high rate

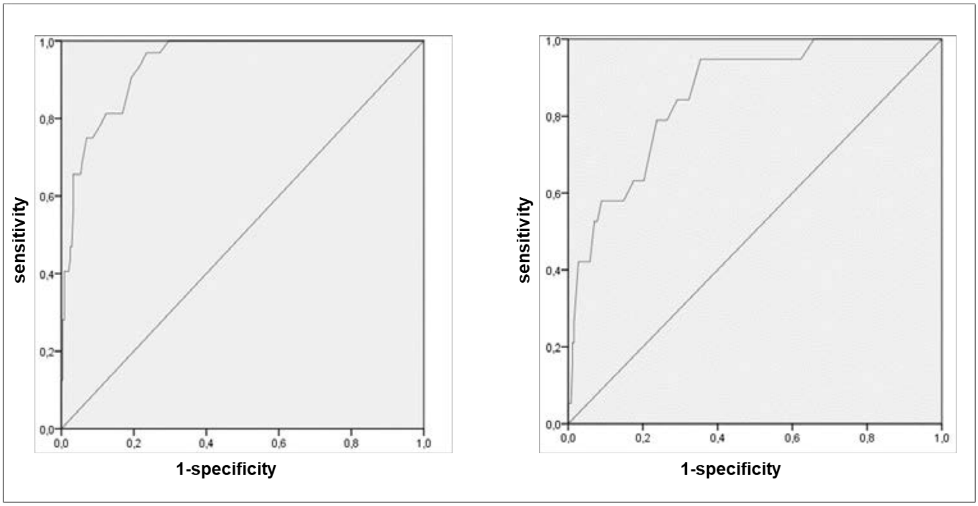


Figure 3
ROC-curves of the game-based screening. The graph on the left displays the curve for the item selection in accordance with M1; the graph on the right displays the curve for the item selection in accordance with M2.

of false positive identification of 16 % vs. 22 %. Both scores produce false negative errors (19 % vs. 6 %). Moreover, with 75 % vs. 91 % we attained high *relative improvement over chance* (RIOC) indices that show how much the hit rate increases compared to the random hit rate.

With respect to the screening assessment of mathematical performance based on M2, the ROC-analysis also revealed positive results. A high AUC value of .86, $p < .001$, 95 % CI [.78, .93] indicated a substantial differentiation capability with the z-scores of -0.63 (percentile rank = 26 %) or -0.24 (percentile rank = 41) as two possible sensitive and specific cut-off values. Using these scores resulted in a sensitivity of 79 % vs. 95 %, a specificity of 76 % vs. 65 %, and false positive identification of 24 % vs. 35 %. The scores showed a 21 % or 5 % failure rate in the identification of low-achieving children. Again, we achieved high RIOC indices of 71 % or 91 %.

Discussion

The study aimed to investigate the applicability, reliability, validity and differentiation capability of a newly developed game-based assessment of early mathematical competencies. The first goal was to explore whether standardized items can be implemented in a game-based assessment tool and be administered while playing a number board game. We examined the number of missing values in the game-based setting as indicators of difficulties in carrying out the game-based assessment. We conducted item analysis and compared children's performance in game-based and standardized assessment settings in order to examine whether the setting affects performance. We found the screening could be conducted, in general, with kindergarten children as young as 4 years of age. Nonetheless, we registered missing data in the game-based setting. The experimenter could decide to skip an item if the child revealed severe performance or motivational problems. Our experimenters had to apply this procedure to only very few cases. Hence, the game-based setting does not seem to affect children in such way that would hinder test application. However, data were also missing in the HoN as the experimenters seemed to have problems in dealing with some of the items in the game-based assessment. This problem affected a comparably high number of assessments. One reason could be the variability in administering each item, as each item was carried out according to standardized instructions but also depended on the progression of the game. The experimenter had to observe the game and decide where and when they could best deliver a certain task within the indicated number range. The observation of the progression of the game parallel to the observation and recording of the children's performance, general adult-child-interaction, and execution of one's own moves in the game meant greater flexibility and higher demands on the experimenter than standardized measurement procedures require. These special demands of game-based settings must be considered in further screening development. On the other hand, difficulties in carrying out the assessment can be clearly assigned to three screening items and seem to originate from unclear instructions or item density. Therefore, the instructions and recording must be amended, or items must be excluded, in order to simplify the process. Item analysis was used to identify tasks with insufficient characteristics, which we had already excluded from data analysis in the present study, and which will be excluded from the screening in future investigations. Additionally, we intend to shorten the procedure with the use of separate screening versions based on one of the developmental models. In summary, item reduction could result in a reduced workload for the user. Moreover, it could also save time in applying the tool. So far, the game-based assessment has clearly comprised fewer items than the standardized measures, but the time required to complete the assessment was almost the same. This is probably because playing the game itself also takes time. Nonetheless, we will have to reduce the time required if we want to develop a resource-saving screening tool.

In the context of applicability, it was also of interest whether children show comparable performance in game-based and standardized measures. Results indicate that 5 and 6-year-old children demonstrated lower performance in the HoN compared to MBK 0. In comparison to MARKO-D, the game produced a slightly higher performance. The differences between game-based and standardized test settings are small but statistically significant. Effect sizes represent small effects in all analyses (Cohen, 1988). Consequently, we cannot finally clarify whether the in-game assessment of mathematical competencies causes performance differences and, if so, whether positive or negative. On the one hand, a reduction in performance could occur because the game-based setting provides multiple cognitive demands. For example, there are many stimuli that must be processed when dealing with the game board, such as rolling the dice, moving on, or drawing event cards. Additionally, children have to switch attention between in-game actions and the numerical tasks, or they might have to inhibit impulsive responses to game-related events. These are critical cognitive processes for academic performance (May et al., 2013). On the other hand, the setting could induce positive or negative affects, which could either enhance or reduce performance (Davis et al., 2007; Graves & Lahey, 1982). Both positive and negative emotional states could arise due to the characteristics of games such as fun, interactivity, engagement, competitiveness, or challenge (Baines & Blatchford, 2011; Hassinger-Das et al., 2017). The results of our study do not clearly demonstrate setting effects in game-based assessment. This is in line with other investigations that were not able to identify an influence of test presentation setting on mathematical achievement (Ricken et al., 2013). With regard to the newly developed screening version of the HoN, further test development and research is needed to investigate whether the game itself is responsible for differences in performance. In this context, it could also be examined whether children perceive game-based and standardized test settings in different ways. Games are considered attractive and engaging for kindergarten children (e.g. Weisberg et al., 2016) which speaks in favor of the use of this approach in early childhood education, but this claim has not yet been investigated systematically.

The second goal of the present study was to evaluate the reliability and validity of the game-based assessment of early mathematical competencies. Results indicate satisfactory internal consistency as well as person and item reliability. Compared to those of standardized measurements (Krajewski, 2018; Ricken et al., 2013; Seeger et al., 2014), reliability is lower. However, since the HoN comprises comparably fewer items, a slight reduction in reliability is expectable.

Consistent with previous studies and the assumptions of both underlying developmental models (Krajewski, 2008; Ricken et al., 2013), results of the present study indicate age-related differences in mathematical competencies assessed with the HoN. This finding can be taken as evidence that the screening reliably captures developmental changes in early mathematical performance. Furthermore, we found strong positive correlations between children's mathematical performance in the HoN and the standardized measurements indicating concurrent validity of the screening tool.

Our third goal was to investigate whether the newly developed game-based screening reliably identifies low-achieving children. We determined cut-off values for low achievement in screening via ROC analyses. In both conditions (M1 and M2), cut-off values could be found, which identified a majority of children who performed below average in standardized measurements. This result indicates a substantial sensitivity for the game-based assessment. ROC-analyses themselves can be seen as quality measures of classifiers and indicated the substantial differentiation capability of the screening tool. High RIOC-indices confirm this finding. According to Marx (1992), RIOC-indices above 66 % are an indication of excellent and specific classifications, which therefore applies to all of the cut-off values considered.

With respect to false negative and false positive classification errors, the first error type can be considered less acceptable in terms of detecting at-risk children. With respect to the prognostic relevance of early mathematical abilities for future fundamental mathematical knowledge (e.g. Geary, 2011; LeFevre et al., 2010; Weißhaupt et al., 2006), it may seem more detrimental to fail to identify low-achieving children via screening, than to give support to some children unnecessarily. This could result in further diagnostics or training being denied to at-risk children. With regard to the present findings, false negative rates in the game-based assessment seem acceptable when choosing the more sensitive cut-off values. In contrast, false positive classification errors could mean further test and training procedures being offered to children who do not actually need them. More sensitive cut-off values are denoted by lower specificity and higher false positive classifications. In the present study, false positive rates for all examined cut-off values are high. This means that 16 % to 35 % of the investigated, yet not affected, children would receive further treatment. The question is whether this implication is reasonable. On the one hand, false positive classifications would provide unnecessary high expenditure for practitioners. On the other hand, further examination or fostering of academic abilities at kindergarten age cannot be considered harmful for children. In addition, developmental screening tools do not serve as differentiated diagnostics. Instead, they function as filters in order to indicate which children should be further examined (Deimann & Kastner-Koller, 2021; Tröster, 2009). Hence, a positive screening result should be dealt with and communicated in a sensible way.

One possible explanation for higher false positive rates could be the proportion of items of low and high difficulty. Since one major aim of the present study was to investigate the applicability of mathematical assessment via a game-based tool in general, we integrated items addressing all of the developmental levels of the underlying models. Similar to other screening tools of mathematical competencies (Seeger et al., 2014), we could consider omitting higher level items in order to increase differentiation in the lower range of performance.

Another possible explanation for high false positive rates could be the effect of higher cognitive requirements of game-based assessment. Processing game activities parallel to mathematical activities might lead to an underestimation of children's actual mathematical performance. As already mentioned, the present findings cannot finally clarify whether in-game assessment affects performance, whether in a positive or negative way. Future studies need to investigate the possible effects of test settings systematically.

Conclusion

In summary, this pilot study is a first step towards the assessment of young children's mathematical competencies by means of a board game. We integrated items based on two different developmental models in a number board game. The double validation of this assessment tool indicated several positive findings. Most of all, the game proved suitable for children as young as four years of age. Traditional test quality measures speak in favour of the reliable and valid assessment for both operationalizations within the new setting. Hence, game-based assessment could be successfully applied across a broad selection of early mathematical competencies. Moreover, the screening detected at least 80 % of low-achieving children.

Nonetheless, the present study also reveals disadvantages of a game-based assessment. This includes the duration of the assessment. Despite an explicit reduction in the number of items, the game lasts as long as the conventional tests do, probably due to the combination of awarding and carrying out each item, and the game action. Moreover, the current version of the game places high demands on the user. Hence, the requirements of the administration and duration of the game clearly need to be reduced. Once these difficulties have been resolved, the administration by practitioners in contrast to highly-trained experimenters also requires examination. Additionally, we need further insights into what game-based assessment means to the participating children. The cognitive demands of in-game assessment and its potential effect on mathematical performance must be systematically investigated. Furthermore, we are interested in examining the motivational, emotional, and social processes of game-based assessment and what these aspects mean for children's engagement and performance. Hence, further research is needed in order to evaluate whether game-based assessment prospectively serves as an attractive addition to conventional tests.

So far, the screening version of the HoN does not exceed the conventional tests in terms of practical manageability, duration, psychometric quality or differentiation ability. Nevertheless, this pilot study shows that we can generally make use of the attractive game-based learning approach as assessment setting in kindergarten age. Hence, we can see the potential of games to enable the integration of diagnostic assessment and training within the same setting, which could strengthen the use of game-based tools in early childhood education.

References

Aragón, E., Cerda, G., Aguilar, M., Mera, C. & Navarro, J. I. (2020). Modulation of general and specific cognitive precursors to early mathematical competencies in preschool children. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-020-00483-4

Aster, M. G. von & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology*, *49*(11), 868–873. https://doi.org/10.1111/j.1469-8749.2007.00868.x

Aunio, P. & Räsänen, P. (2016). Core numerical skills for learning mathematics in children aged five to eight years – a working model for educators. *European Early Childhood Education Research Journal*, *24*(5), 684–704. https://doi.org/10.1080/1350293X.2014.996424

Aunola, K., Leskinen, E., Lerkkanen, M. K. & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, *96*(4), 699–713. https://doi.org/10.1037/0022-0663.96.4.699

Baines, E. & Blatchford, P. (2011). Children's games and playground activities in school and their role in development. In A. D. Pellegrini (Hrsg.), *The oxford handbook of the development of play* (pp. 260–283). Oxford University Press.

Bayeck, R. Y. (2020). Examining board gameplay and learning: A multidisciplinary review of recent research. *Simulation & Gaming*, *51*(4), 411–431. https://doi.org/10.1177/1046878119901286

Berner, V.-D., Seitz-Stein, K., Segerer, R., Oesterlen, E. & Niklas, F. (2021). "Good" or "Well Calculated"? Effects of feedback on performance and self-concept of 5- to 7-year-old children in math. *Educational Psychology*. https://doi.org/10.1080/01443410.2021.2001790

Bock-Famulla, K., Lange, J. & Strunz, E. (2015). *Länderreport Frühkindliche Bildungssysteme 2015 [Report on early childhood education systems 2015]*. Bertelsmann Stiftung.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C. & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Chu, F. W., vanMarle, K. & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology*, *132*, 205–212. https://doi.org/10.1016/j.jecp.2015.01.006

Chu, F. W., vanMarle, K., Rouder, J. & Geary, D. C. (2018). Children's early understanding of number predicts their later problem-solving sophistication in addition. *Journal of Experimental Child Psychology*, *169*, 73–92. https://doi.org/10.1016/j.jecp.2017.12.010

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

Costa, H. M., Nicholson, B., Donlan, C. & van Herwegen, J. (2018). Low performance on mathematical tasks in preschoolers: the importance of domain-general and domain-specific abilities. *Journal of Intellectual Disability Research*, *62*(4), 292–302. https://doi.org/10.1111/jir.12465

Cueli, M., Areces, D., McCaskey, U., Álvarez-García, D. & González-Castro, P. (2019). Mathematics competence level: The contribution of non-symbolic and spatial magnitude comparison skills. *Frontiers in psychology*, *10*, 465. https://doi.org/10.3389/fpsyg.2019.00465

Davis, M. A., Kirby, S. L. & Curtis, M. B. (2007). The influence of affect on goal choice and task performance. *Journal of Applied Social Psychology*, *37*(1), 14–42.

Deimann, P. & Kastner-Koller, U. (2021). Entwicklungsscreening [Developmental screening]. In M. A. Wirtz (Hrsg.), *Dorsch – Lexikon der Psychologie*. https://dorsch.hogrefe.com/stichwort/entwicklungsscreening

Ehlert, A., Ricken, G. & Fritz, A. (2020). *Mathematik- und Rechenkonzepte im Vorschulalter – Screening [Assessment of math concepts in preschool aged children - screening tool]*. Hogrefe.

Ehm, J. H., Lonnemann, J. & Hasselhorn, M. (2017). *Wie Kinder zwischen vier und acht Jahren lernen. Psychologische Erkenntnisse und Konsequenzen für die Praxis [How children learn between four and eight years of age. Insights and consequences for practice]*. Kohlhammer.

Elofsson, J., Gustafson, S., Samuelsson, J. & Träff, U. (2016). Playing number board games supports 5-year-old children's early mathematical development. *The Journal of Mathematical Behavior*, *43*, 134–147. https://doi.org/10.1016/j.jmathb.2016.07.003

Feigenson, L., Dehaene, S. & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Ferreira, F. d. O., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K. & Haase, V. G. (2012). Explaining school mathematics performance from symbolic and nonsymbolic magnitude processing: Similarities and differences between typical and low-achieving children. *Psychology & Neuroscience*, *5*(1), 37–46. https://doi.org/10.3922/j.psns.2012.1.06

Fischer, U., Roesch, S. & Moeller, K. (2017). Diagnostik und Förderung bei Rechenschwäche. Messen wir, was wir fördern wollen? [Dyscalculia diagnosis and treatment. Are we measuring what we want to improve?]. *Lernen und Lernstörungen*, *6*(1), 25–38. https://doi.org/10.1024/2235-0977/a000160

Fritz, A., Ehlert, A. & Leutner, D. (2018). Arithmetische Konzepte aus kognitiv-entwicklungspsychologischer Sicht [Arithmetic concepts from a cognitive developmentalpsychology perspective]. *Journal für Mathematik-Didaktik*, *39*(1), 7–41. https://doi.org/10.1007/s13138-018-0131-6

Fritz, A. & Ricken, G. (2008). *Rechenschwäche [Acalculia]*. Reinhardt.

Garon-Carrier, G., Boivin, M., Lemelin, J. P., Kovas, Y., Parent, S., Séguin, J. R., Vitaro, F., Tremblay, R. E. & Dionne, G. (2018). Early developmental trajectories of number knowledge and math achievement from 4 to 10 years: Low-persistent profile and early-life predictors. *Journal of School Psychology*, *68*, 84–98. https://doi.org/10.1016/j.jsp.2018.02.004

Gasteiger, H. (2013). Förderung elementarer mathematischer Kompetenzen durch Würfelspiele - Ergebnisse einer Interventionsstudie [Education of basic mathematical competencies by means of dice games - Results of an intervention study]. In G. Greefrath, F. Käpnick & M. Stein (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 336–339). WTM.

Gasteiger, H. & Moeller, K. (2021). Fostering early numerical competencies by playing conventional board games. *Journal of Experimental Child Psychology*, *204*, 105060. https://doi.org/10.1016/j.jecp.2020.105060

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539–1552. https://doi.org/10.1037/a0025510

Gerster, H. D. (2003). Schwierigkeiten beim Erwerb arithmetischer Konzepte im Anfangsunterricht [Learning difficulties in arithmetical concepts at primary school]. In F. Lenart, N. Holzer & H. Schaupp (Hrsg.), *Rechenschwäche, Rechenstörung, Dyskalkulie: Erkennung, Prävention, Förderung* (S. 154–160). Leykam.

Graves, K. J. & Lahey, B. B. (1982). The effects of induced affective states on letter discrimination in disadvantaged preschool children. *Journal of Applied Developmental Psychology*, *3*(2), 149–154. https://doi.org/10.1016/0193-3973(82)90025-9

Häsel-Weide, U., Nührenbörger, M., Moser Opitz, E. & Wittich, C. (2014). *Ablösung vom zählenden Rechnen. Fördereinheiten für heterogene Lerngruppen [Overcoming counting principles in arithmetic. Tasks samples for heterogeneous learning groups]*. Klett Kallmeyer.

Hassinger-Das, B., Toub, T. S., Zosh, J. M., Michnick, J., Golinkoff, R. & Hirsh-Pasek, K. (2017). More than just fun: A place for games in playful learning / Más que diversión: El lugar de los juegos reglados en el aprendizaje lúdico. *Infancia Y Aprendizaje: Journal for the Study of Education and Development*, *40*(2), 191–218. https://doi.org/10.1080/02103702.2017.1292684

Hauser, B., Vogt, F., Stebler, R. & Rechsteiner, K. (2014). Förderung früher mathematischer Kompetenzen. Spielintegriert oder trainingsbasiert [Promoting early mathematical competences. Play-integrated or training-based]. *Frühe Bildung*, *3*(3), 139–145. https://doi.org/10.1026/2191-9186/a000144

Henning, E., Ehlert, A., Balzer, L., Ragpot, L., Herholdt, R. & Fritz, A. (2019). *MARKO-D South Africa. Assessment of number concept development (Afrikaans, English, IsiZulu, and Sesotho version)*. University of Johannesburg.

Herzog, M. & Fritz, A. (2022). Entwicklungs- und spielbasierte Förderung früher arithmetischer Konzepte [Developmental and game-based intervention for early arithmetic concepts]. *Lernen und Lernstörungen*, *11*(1), 5–17. https://doi.org/10.1024/2235-0977/a000358

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E. & Singer, D. G. (2009). *A mandate for playful learning in school: Presenting the evidence*. Oxford University Press.

Jörns, C., Schuchardt, K., Grube, D., Barkam, L. & Mähler, C. (2017). ZIKZAK – Profitieren Kindergartenkinder von Gesellschaftsspielen zur Förderung numerischer Kompetenzen und phonologischer Bewusstheit? [ZIKZAK – A game-orientated training to boost numerical skills and phonological awareness at preschool age]. *Unterrichtswissenschaft*, *45*(3), 220–238.

Jörns, C., Schuchardt, K., Grube, D. & Mähler, C. (2014). Spielorientierte Förderung numerischer Kompetenzen im Vorschulalter und deren Eignung zur Prävention von Rechenschwierigeiten [Playful training of numerical skills in preschool for prevention of mathematical disability]. *Empirische Sonderpädagogik*, *3*, 243–259.

Jörns, C., Schuchardt, K., Mähler, C. & Grube, D. (2013). Alltagsintegrierte Förderung numerischer Kompetenzen imKindergarten [Possible improvement of mathematical competencies in kindergarten]. *Frühe Bildung*, *2*(2), 84–91.

Kiili, K., Moeller, K. & Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators. *Computers & Education*, *120*, 13–28. https://doi.org/10.1016/j.compedu.2018.01.012

Krajewski, K. (2008). *Vorhersage von Rechenschwäche in der Grundschule [Prediction of mathematics achievement deficiencies in elementary school students]*. Kovac.

Krajewski, K. (2014). Test zur Erfassung mathematischer Basiskompetenzen im Kindergartenalter (MBK-0) [Test of basic mathematical competencies in preschool children]. In D. Seeger, M. Holodynski & E. Souvignier (Hrsg.), *BIKO 3-6. BIKO-Screening zur Entwicklung von Basiskompetenzen für 3- bis 6-Jährige*. Hogrefe.

Krajewski, K. (2018). *MBK 0: Test mathematischer Basiskompetenzen im Kindergartenalter [Test of basic mathematical competencies in preschool children]*. Hogrefe.

Krajewski, K. & Schneider, W. (2006). Mathematische Vorläuferfertigkeiten im Vorschulalter und ihre Vorhersagekraft für die Mathematikleistung bis zum Ende der Grundschulzeit [Predicting mathematical school achievement by mathematical precursors assessed in kindergarten]. *Psychologie in Erziehung und Unterricht*, *52*, 246–262.

Krajewski, K. & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology*, *103*(4), 516–531. https://doi.org/10.1016/j.jecp.2009.03.009

Lange, A. A., Brenneman, K. & Sareh, N. (2021). Using number games to support mathematical learning in preschool and home environments. *Early Education and Development*, *32*(3), 459–479. https://doi.org/10.1080/10409289.2020.1778386

Laski, E. V. & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology*, *50*(3), 853–864. https://doi.org/10.1037/a0034321

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D. & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Lenhard, W. & Lenhard, A. (2014). *Calculation of test quality criteria for screenings*. http://www.psychometrica.de/testkennwerte.html

Lenhard, W. & Lenhard, A. (2016). *Berechnung von Effektstärken [Calculation of effect sizes]*. https://www.psychometrica.de/effect_size.html

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D. & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR serious games*, *4*(2). https://doi.org/10.2196/games.5888

Martin, R., Cirino, P. T., Sharp, C. & Barnes, M. (2014). Number and counting skills in kindergarten as predictors of grade 1 mathematical skills. *Learning and Individual Differences*, *34*, 12–23. https://doi.org/10.1016/j.lindif.2014.05.006

Marx, H. (1992). *Vorhersage von Lese-Rechtschreibschwierigkeiten in Theorie und Anwendung [Prediction of reading and spelling difficulties in theory and application]* [Unpublished habilitation thesis]. Bielefeld University, Bielefeld.

May, T., Rinehart, N., Wilding, J. & Cornish, K. (2013). The role of attention in the academic attainment of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*(9), 2147–2158. https://doi.org/10.1007/s10803-013-1766-2

Merkley, R. & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, *10*, 14–20. https://doi.org/10.1016/j.cobeha.2016.04.006

Navarro, J. I., Aguilar, M., Marchena, E., Ruiz, G., Menacho, I. & van Luit, J. E. H. (2012). Longitudinal study of low and high achievers in early mathematics. *The British Journal of Educational Psychology*, *82*(1), 28–41. https://doi.org/10.1111/j.2044-8279.2011.02043.x

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C. & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly*, *36*, 550–560. https://doi.org/10.1016/j.ecresq.2016.02.003

Ninaus, M., Kiili, K., McMullen, J. & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*, *70*, 197–206. https://doi.org/10.1016/j.chb.2017.01.004

Ninaus, M., Pereira, G., Stefitz, R., Prada, R., Paiva, A., Neuper, C. & Wood, G. (2015). Game elements improve performance in a working memory training task. *International Journal of Serious Games*, *2*(1). https://doi.org/10.17083/ijsg.v2i1.60

Ponticorvo, M., Schembri, M. & Miglino, O. (2021). Training and assessing numerical abilities across the lifespan with intelligent systems: The example of Baldo. *Expert Systems*, *39*(1). https://doi.org/10.1111/exsy.12817

Ramani, G. B. & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*(2), 375–394. https://doi.org/10.1111/j.1467-8624.2007.01131.x

Ramani, G. B., Siegler, R. S. & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology*, *104*(3), 661–672. https://doi.org/10.1037/a0028995

Ricken, G., Fritz, A. & Balzer, L. (2013). *MARKO-D: Mathematik- und Rechenkonzepte im Vorschulalter - Diagnose [Assessment of math concepts in pre-school aged children]*. Hogrefe.

Sasanguie, D., Smedt, B. de, Defever, E. & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *The British Journal of Developmental Psychology*, *30*(Pt 2), 344–357. https://doi.org/10.1111/j.2044-835X.2011.02048.x

Scalise, N. R., Daubert, E. N. & Ramani, G. B. (2020). Benefits of playing numerical card games on head start children's mathematical skills. *The Journal of Experimental Education*, *88*(2), 200–220. https://doi.org/10.1080/00220973.2019.1581721

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J. & Smedt, B. de (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, *20*(3). https://doi.org/10.1111/desc.12372

Seeger, D., Holodynski, M. & Souvignier, E. (2014). *BIKO 3-6. BIKO-Screening zur Entwicklung von Basiskompetenzen für 3- bis 6-Jährige [BIKO-Screening for the development of basic skills of 3 to 6 year old children]*. Hogrefe.

Shih, S.-C., Kuo, B.-C. & Lee, S.-J. (2019). An online game-based computational estimation assessment combining cognitive diagnostic model and strategy analysis. *Educational Psychology*, *39*(10), 1255–1277. https://doi.org/10.1080/01443410.2018.1501468

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C.-P., Kuba, R., Liu, Z., Yang, X. & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, *37*(1), 127–141. https://doi.org/10.1111/jcal.12473

Siegler, R. S. & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*(3), 545–560. https://doi.org/10.1037/a0014239

Skillen, J., Berner, V.-D., Ricken, G. & Seitz-Stein, K. (2018). Förderung basisnumerischer Kompetenzen mit einem regelbasierten Spiel im Einzel- und Gruppensetting [Education of basic numerical competencies with a game in one-to-one and group settings]. *Frühe Bildung*, *7*(3), 144–151. https://doi.org/10.1026/2191-9186/a000381

Skillen, J., Berner, V.-D. & Seitz-Stein, K. (2018). The rule counts! Acquisition of mathematical competencies with a number board game. *The Journal of Educational Research*, *111*(5), 554–563. https://doi.org/10.1080/00220671.2017.1313187

Textor, M. R. (2005). *Bildung: Was müssen und können Kindergärten leisten? [Education: What kindergartens have to cope with]*. https://www.kindergartenpaedagogik.de/fachartikel/bildung-erziehung-betreuung/1230

Tröster, H. (2009). *Früherkennung im Kindes- und Jugendalter: Strategien bei Entwicklungs-, Lern- und Verhaltensstörungen [Early detection in childhood and adolescence]*. Hogrefe.

van Vuuren, E., Herzog, M. & Fritz, A. (2021). Theory-based early numeracy programme adapted to learners' pre-knowledge: The 'Meerkat Maths' programme. In S. Gravett & E. Henning (Hrsg.), *Glimpses into primary school teacher education in South Africa*. Routledge.

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K. & Klahr, D. (2016). Guided play: Principles and practices. *Current Directions in Psychological Science*, *25*(3), 177–182. https://doi.org/10.1177/0963721416645512

Weißhaupt, S., Peucker, S. & Wirtz, M. (2006). Diagnose mathematischen Vorwissens im Vorschulalter und Vorhersage von Rechenleistungen und Rechenschwierigkeiten in der Grundschule [Measuring preschool mathematical knowledge and predicting students' abilities and deficiencies in mathematics in primary school]. *Psychologie in Erziehung und Unterricht*, *53*, 236–245.

Whyte, J. C. & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology*, *44*, 588–596. https://doi.org/10.1037/0012-1649.44.2.588

Wittmann, E. C. & Müller, G. N. (1995). *Handbuch produktiver Rechenübungen. Vom Einspluseins zum Einmaleins [Manual of productive arithmetic tasks. From addition to multiplication]*. Klett.

Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.

Autorinnen- und Autorenhinweis:

Johanna Skillen
https://orcid.org/0000-0001-5683-0247

Gabi Ricken
https://orcid.org/0000-0003-3845-2873

We would like to thank all the children, parents, and educators who participated in this research. Special thanks goes to Nadine Adolf, Tanja Bannehr, Kevin Beesk, Katharina Betz, Johanna Beyschlag, Sophie Bösl, Veronika Dexl, Daniela Dittenhauser, Laura Fumy, Anne Hungbaur, Romina Luderer, Frank Palka, Katharina Humpmayer, Anja Steinmüller, Lisa Tomm and Christine Vietor who assisted with data collection.

Korrespondenzadresse:
Johanna Skillen
Lehrstuhl für Entwicklungs- und
Pädagogische Psychologie
Katholische Universität Eichstätt-Ingolstadt
Ostenstraße 25,
85072 Eichstätt
Johanna.Skillen@ku.de

Erstmals eingereicht: 26.07.2022
Überarbeitung eingereicht: 30.08.2022
Angenommen: 21.01.2023

Offene Daten	Keine Angabe
Offener Code	Keine Angabe
Offene Materialien	Die Materialien des spielbasierten Screenings können bei den Autorinnen angefragt werden. Die Materialien der standardisierten Verfahren sind veröffentlicht und über den entsprechenden Verlag beziehbar.
Präregistrierung	Nein
Votum Ethikkommission	Nein. Die Studie wurde von den teilnehmenden Kindergärten und ihren Trägern geprüft und genehmigt.
Finanzielle und weitere sachliche Unterstützung	Keine
Autorenschaft	Idee und Konzeption: JS und KSS, Planung und Durchführung: JS, Datenanalyse: JS, Verschriftlichung: JS, KSS und GR