Early Childhood Research Quarterly 54 (2021) 164–178

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Enhancing children's numerical skills through a play-based intervention at kindergarten and at home: a quasi-experimental study

Anne-Françoise de Chambrier[a,*], Ariane Baye[b], Mélanie Tinnes-Vigne[c], Youssef Tazouti[d], Joëlle Vlassis[c], Débora Poncelet[c], Nadine Giauque[a], Annick Fagnant[b], Christophe Luxembourger[d], Amélie Auquière[b], Sylvie Kerger[c], Christophe Dierendonck[c]

[a] University of Teacher Education from State of Vaud, Avenue de Cour 33, 1014 Lausanne, Switzerland
[b] University of Liege, Quai Roosevelt 1b, 4000 Liège, Belgium
[c] University of Luxembourg, Maison des Sciences Humaines 11, Porte des Sciences, L-4366 Esch-sur-Alzette, Grand Duchy of Luxembourg
[d] University of Lorraine, 91 avenue de la Libération, BP 32142, 54021 Nancy Cedex, France

ARTICLE INFO

Article history:
Received 21 May 2019
Received in revised form 22 August 2020
Accepted 5 September 2020

Keywords:
Early number skills
Kindergarten
Play-Based intervention
Parental involvement
Multilevel analyses

ABSTRACT

Early number skills are critical predictors of academic achievement, which is why focusing on their instruction from the very beginning of education is recommended. Young children's number knowledge is also strongly influenced by home numeracy practices. This 12-week quasi-experimental study tested whether early number skills could be enhanced by a play-based intervention implemented at kindergarten (aged 4–6 years) by the teachers and whether providing numerical games to families delivered added value. A total of 569 children from 46 kindergarten classes were assigned either to the first treatment condition in which games were played only at kindergarten, or to the second treatment condition in which games were played at kindergarten and at home, or to the business-as-usual control condition. Measures of numerical ability were collected at pretest and posttest and analyzed through item response theory and multilevel modeling. Results indicated that playing the games at kindergarten allowed children with average and above-average initial performance to make more progress than children in the control group, while providing the games at home allowed low achievers from various backgrounds to progress more than in the other conditions. Implications for early mathematics instruction and for home-based intervention studies are discussed.

1. Introduction

Early number skills have received growing attention in the past decade since they are good predictors of later mathematical skills, general academic achievement, and quality of life (Every Child a Chance Trust, 2009). Recent longitudinal studies have shown that early numerical skills predict mathematical achievement from first grade to high school and are positively related to socioeconomic status at middle life (Duncan et al., 2007; Jordan, Glutting, & Ramineni, 2009; Ritchie & Bates, 2013). Despite this converging evidence, there exists far less research concerning how to support early numerical skills compared to literacy, which contributes to perpetuating the lesser emphasis placed on numeracy in the lives of young children (Mazzocco & Claessens, 2020).

Early number skills refer to a whole set of elementary competencies comprising oral counting, enumerating, number relations, collection comparison, arithmetic counting strategies, and number decomposition. For example, to enumerate collections, children must master the number-word sequence to be able to match the number words to the items to be counted and to know that the last number in the count denotes the whole number of objects (Fuson, Richard, & Briars, 1982; Gelmann & Gallistel, 1978). Children must also be aware of order relations between numbers to accurately compare and classify quantities. Progressively, young children learn to solve basic arithmetic problems. They generally start by counting concrete objects, move to counting number words, and finally use mental strategies such as reasoning or recall (Baroody, 1987). Another key concept is understanding that numbers can be broken into subsets, referred to as number decom-

---

* Corresponding author.
E-mail addresses: anne-francoise.de-chambrier@hepl.ch (A.-F. de Chambrier), Ariane.Baye@uliege.be (A. Baye), melanie.tinnes-vigne@uni.lu (M. Tinnes-Vigne), youssef.tazouti@univ-lorraine.fr (Y. Tazouti), joelle.vlassis@uni.lu (J. Vlassis), debora.poncelet@uni.lu (D. Poncelet), nadine.giauque@hepl.ch (N. Giauque), aFagnant@uliege.be (A. Fagnant), christophe.luxembourger@univ-lorraine.fr (C. Luxembourger), amelie.auquiere@uliege.be (A. Auquière), sylvie.kerger@uni.lu (S. Kerger), christophe.dierendonck@uni.lu (C. Dierendonck).

https://doi.org/10.1016/j.ecresq.2020.09.003

position (Dyson, Jordan, & Glutting, 2013). These early number skills have been shown to be developmental precursors of more advanced formal concepts (National Mathematics Advisory Panel, 2008, NMAP, 2008; National Research Council, 2009, NRC, 2009).

1.1. Early number skills at kindergarten

Given the importance of early number skills, researchers and national commissions have emphasized the necessity of focusing upon early mathematics education (Clements & Sarama, 2007; Frye et al., 2013; Ginsburg, Lee, & Boyd, 2008; National Council of Teachers of Mathematics, 2006, NCTM, 2006; NMAP, 2008; NRC, 2009; Starkey, Klein, & Wakeley, 2004). Prekindergarten and kindergarten (aged 4–6 years) are widely acknowledged to be an especially important time for preparing pupils for success in math (Clements & Sarama, 2007). These early education years represent an opportunity to narrow the gaps in attainment between children entering with varied number knowledge (Ginsburg et al., 2008; Scalise, Daubert, & Ramani, 2017). Indeed, children who leave kindergarten with weak number skills may never catch up to those who enter first grade with good number competencies, whereas early interventions can place at-risk children onto an appropriate learning trajectory (Jordan, Kaplan, Ramineni, & Locuniak, 2010). However, data has demonstrated that many preschool and kindergarten teachers consider math less important than other acquisition domains. These teachers tend to think that everyday situations or a general enrichment of the classroom are sufficient, or even believe that mathematics is too difficult for young children (Cannon & Ginsburg, 2008; Hachey, 2013; Lee & Ginsburg, 2009; Platas, 2014). Instead, it is nowadays recognized that early number skills must be taught intentionally from the very beginning of education (Cannon & Ginsburg, 2008; Clements & Sarama, 2011; Park, Bermudez, Roberts, & Brannon, 2016). In this context, teachers are expected to delineate teaching objectives and to create opportunities for children to acquire important early numerical competencies.

Recently, then, many mathematic interventions have been elaborated to support the development of early mathematical skills. These interventions differ greatly in their approaches towards targeted competencies, duration of training, targeted children, type of intervention, nature of the material, and performance assessment. Some interventions have targeted a wide set of mathematical competencies (e.g., Clements & Sarama, 2007), and others have focused on one or two specific numerical skills (e.g., Whyte & Bull, 2008). The duration of interventions has varied from very few and short sessions (e.g., Ramani & Siegler, 2011) to several months or years (e.g., Lewis Presser, Clements, Ginsburg, & Ertle, 2015). Some interventions have been implemented in the whole class and/or small groups of children (e.g., Clarke et al., 2011), while others have been administered individually, often outside of the classroom (e.g., Scalise et al., 2017). A few interventions have been designed for children with a wide spectrum of skills (e.g., Codding, Chan-Iannetta, George, Ferreira, & Volpe, 2011), whereas several interventions have specifically addressed low-performing or at-risk children (e.g., Aunio & Mononen, 2017).

Some interventions have been formal, involving planned lessons organized into units designed to be led by an adult and sometimes linked to educational standards (e.g., Dyson et al., 2013), whilst others have been more informal, using games (e.g., Whyte & Bull, 2008). Notably, interventions have mostly been implemented by a research assistant (e.g., Friso-van den Bos, Kroesbergen, & Van Luit, 2018) and rarely by the children's teachers (e.g., Starkey et al., 2004). Some interventions have required particular materials such as computers (e.g., Baroody, Eiland, Purpura, & Reid, 2012), whereas others have made use of more affordable materials (e.g., Bojorque, Torbeyns, Van Hoof, Van Nijlen, & Verschaffel, 2018). Finally, most of the researchers have employed standardized and validated batteries of tests to measure the children's skills (e.g., Clements, Sarama, Spitler, Lange, & Wolfe, 2011), but some have used researcher-designed tasks that covered the precise content of the intervention (i.e., Ramani & Siegler, 2008). For example, Lewis Presser et al. (2015) measured the effects of a 2-year implementation of the *Big Math For Little Kids* program. This curriculum was designed for children aged 4 and 5 years and targeted many domains of mathematics (number, shape, patterns and logic, measurement, number operations, and spatial relations). The lessons were conducted by the classroom teacher for 20−30 min per day and addressed all the children in the classroom through group instruction, small-group teaching, and individual exploration, which mainly took the form of games and activities with manipulatives and stories. The children's skills were assessed by a whole validated battery not aligned with the curriculum. By contrast, Siegler and Ramani's (2008) intervention comprised four 15-minute sessions conducted with low-income children individually met by an experimenter outside of the classroom to improve their ability to place numbers from 1 to 10 on a number line through a board game. Skills were assessed by a researcher-made task that fully aligned with the content of the intervention.

Furthermore, several of the studies that specifically addressed low-performing or at-risk children have demonstrated that their numerical skills could also be significantly enhanced relative to a control group (Dyson et al., 2013; Kroesbergen & Van Luit, 2003; Scalise et al., 2017; Schacter & Jo, 2016; Shanley, Clarke, Doabler, Kurtz-Nelson, & Fien, 2017; Siegler & Ramani, 2008; Sood & Jitendra, 2011; Van Luit & Shopman, 2000). These children's gains were sometimes even higher than the gains of their not-at-risk peers (Clarke et al., 2016; Ramani & Siegler, 2011).

A recent meta-analysis conducted on the effectiveness of (pre)kindergarten mathematical interventions showed that the average effect size for all programs was moderate (d = 0.62), indicating that interventions designed to improve young children's mathematical skills can clearly have an effect (Wang, Firmender, Power, & Byrnes, 2016). This gives good reason to continue to develop such intervention studies. However, the many significant differences between the various interventions make these studies difficult to compare. The moderator analysis conducted in Wang et al.'s meta-analysis (2016) allowed to begin to identify some of the characteristics of interventions that give rise to larger effects, such as providing the interventions individually and intensively, targeting specific content, and assessing skills through researcher-made tasks. Nevertheless, additional intervention studies of different types are needed to better identify the instructional components that render such programs more effective (Frye et al., 2013). More precisely, interventions that can be qualified as "ecologically valid" (i.e., conducted by the teacher with all the children of the class, targeting several skills) are scarce and knowing under which circumstances such interventions might be effective for the different children of a classroom is essential.

1.2. Early number skills at home

Although teaching mathematics at preschool and kindergarten is critical, the home environment is the first influence on child numeracy (LeFevre et al., 2010; Niklas & Schneider, 2014). Children's knowledge of numbers has been shown to be positively associated with the frequency of parent-child numeracy activities, such as playing counting games, talking about numbers, or providing direct instruction about numbers (Hill & Craft, 2003; Kleemans, Peeters, Segers, & Verhoeven, 2012; LeFevre et al., 2009; Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010). Home numeracy practices greatly differ between families, such activities tending to be rarer and poorer in low-income families

(Ramani, Rowe, Eason, & Leech, 2015; Vandermaas-Peeler, Nelson, Bumpass, & Sassine, 2009). For example, it has been found that disadvantaged parents often fail to grasp opportunities to initiate number-related exchanges with their children or tend to limit this support to counting (Anderson, 1997; Cheung & McBride, 2017; Skwarchuk, 2009; Zhou et al., 2006). Although parental involvement can be expressed in many ways that are not always perceived by the academic institution (Payet, 2017), many parents themselves report they do not know the numeracy concepts that their young children should learn nor how to support them (Cannon & Ginsburg, 2008; Skwarchuk, 2009). Furthermore, they effectively appear to be less equipped to support their children's numeracy development compared with their literacy skills (Niklas, Cohrssen, & Tayler, 2016). The negative feelings that many parents hold towards the subject of mathematics have also been suggested as an explanation for why they tend to neglect the domain (Cannon & Ginsburg, 2008).

Given the importance of home numeracy practices to children's learning, recent studies have implemented family-based mathematical interventions (Cheung & McBride-Chang, 2015, 2017; Niklas et al., 2016; Sonnenschein, Metzger, Dowling, & Simons, 2016; Ramani & Scalise, 2020; Starkey & Klein, 2000; Starkey et al., 2004; Vandermaas-Peeler, Ferretti, & Loving, 2012). Home-based interventions have been implemented more often for early literacy, and the results have shown that providing benchmarks to parents concerning what could be accomplished at home improved their children's skills and knowledge (Senechal & Young, 2008).

Some studies have also measured the effects of numeracy interventions implemented at home (Cheung & McBride-Chang, 2015; Cheung & McBride, 2017; Dulay, Cheung, Reyes, & McBride, 2018; Niklas et al., 2016; Ramani & Scalise, 2020; Sonnenschein et al., 2016; Starkey & Klein, 2000; Vandermaas-Peeler et al., 2012). In most of these studies, parents were provided with one or two games and were instructed how to foster some specific numerical skills by using them (Cheung & McBride-Chang, 2015; Cheung & McBride, 2017; Dulay et al., 2018; Niklas et al., 2016; Sonnenschein et al., 2016; Vandermaas-Peeler et al., 2012). Most of these studies have highlighted that the numerical skills of the children who benefitted from an intervention at home improved significantly more than those of children in the control group (Cheung & McBride, 2015; Dulay et al., 2018; Vandermaas-Peeler et al., 2012), and this was also the case when targeting low-performing/at-risk children (Starkey & Klein, 2000).

To our knowledge, only two studies have measured the impact of an intervention implemented at prekindergarten *and* at home relative to a business-as-usual condition (Klein, Starkey, Clements, Sarama, & Iyer, 2008; Starkey et al., 2004). These two studies implemented a prekindergarten mathematical curriculum in the classes and included the parents through parallel activities. They targeted children from low- to middle-income families (Starkey et al., 2004) or disadvantaged children (Klein et al., 2008). They found that children in the experimental group progressed significantly more than their peers in the control group. Starkey et al. (2004) even found that relative to their starting point, low-income children acquired more knowledge than middle-income children.

According to our review of the literature, only one study published in French has measured the added value of having parents participate in a kindergarten intervention relative to a mathematical intervention implemented only at kindergarten (Jalbert & Pagani, 2007). In this study (Rightstart program), parents participated in four training workshops, half of them with their child. Through hierarchical regression analyses, the researchers found that the children whose parents were trained progressed slightly more than the children who benefitted from the intervention implemented only at kindergarten. Thus, to the best of our knowledge, this is the only study to demonstrate that involving parents in a kindergarten intervention can be more effective than fostering the children's number skills at kindergarten only. More studies of this type and deeper analyses would allow to know whether the added value of parental involvement is true for children of different levels of skills and to identify types of interventions likely to enroll parents from different backgrounds.

1.3. A play-based intervention

Several sources have suggested that a play-based intervention – rather than a more formal intervention – could be particularly suitable both for the (pre)kindergarten and the home environment. Data suggests that playful learning activities lead to greater learning than non-playful learning activities, particularly for young children (Ferrara, Hirsh-Pasek, Newcombe, Golinkoff, & Lam, 2011; Hirsh-Pasek, Golinkoff, Berk, & Singer, 2009; Scalise et al., 2017). Playful numerical activities also encourage young children's interest in mathematics more than formal numerical activities (Cheung & McBride, 2017). In their recent review of the literature, Hirsh-Pasek et al. (2015) highlighted how the four components that emerged as fostering better learning conditions (i.e., children being active, engaged, and interacting with the material and with peers) were all inherent characteristics of games. Moreover, numerical games could better convince skeptical kindergarten teachers that important numerical concepts are entirely within the reach of young children. Playing numerical games might also help parents who have negative feelings toward mathematics to assist their child's numeracy development in a pleasant manner. Notably, emphasizing enjoyment rather than educational value has been suggested to be more likely to help with the enrollment of low-income families in home-based interventions (Lingwood, Billington, & Rowland, 2018). Additionally, playing internationally known games familiar to many parents, such as dominos, bridge cards, and battle or board games, might contribute to bridging the gap between schools and families from many parts of the world.

1.4. The current study

In view of the above, it is crucial to enhance young children's numerical skills both at kindergarten and at home (Anders et al., 2012). Based on the literature, some characteristics of such an intervention might be particularly relevant. First, a play-based intervention targeting a broad range of number skills – rather than one or two specific number skills – could provide both a turnkey tool to teachers and benchmarks for parents regarding the main early number skills to be developed. Furthermore, measuring the added value of including parents in such a program compared with an intervention implemented only at kindergarten is essential. Indeed, it is worth knowing whether enrolling parents might enhance children's number skills beyond levels achieved by the school environment alone. Notably, such studies have very rarely been conducted (Jalbert & Pagani, 2007).

Another aspect that is critical when studying the potential of an intervention is its replicability and its sustainability in real-life school conditions. Indeed, an intervention found to be effective but that cannot be pursued when the research grants are no longer available may never again be conducted. These concerns of replicability and sustainability apply to both the financial resources available to schools and the teachers' professional reality. In addition to ensuring its feasibility, measuring the effect of an intervention conducted by the teacher within the classroom allows to test the effectiveness of such realistic implementation conditions (Slavin & Lake, 2008). It also measures the effectiveness of an intervention without inflating its effect size, compared with what occurs when treatments are delivered by researchers (Baye, Inns, Lake, & Slavin, 2018). Assessing the effect of an intervention implemented by the classroom teacher in a class-wide setting, and using inexpensive materials, is thereby likely to provide very relevant information to school stakeholders.

Conducted in four countries across continental Europe (Belgium, France, Luxembourg, and Switzerland), this quasi-experimental study aimed to measure the efficacy of a 12-week numerical games intervention implemented by trained teachers in whole-class groups intended to foster a broad range of basic numerical skills in kindergartners (attending the two years before primary school, children aged 4–6 years). Games were proposed in two experimental conditions: at kindergarten and both at kindergarten and in families. We first hypothesized that children attending classrooms that implemented the intervention would show higher gains in numerical skills compared with children in the business-as-usual control classrooms. Second, because of the expected impact of home numeracy practices, we hypothesized that children playing at home, in addition to playing at kindergarten, would demonstrate greater progress compared with children playing only at kindergarten and with children in the control group. Third, by targeting basic number skills to narrow the gaps between low and high achievers, we anticipated that in both experimental conditions the intervention's effects would be greater for pupils with low initial skills, as found in several previous studies (Clarke et al., 2011, 2016; Ramani & Siegler, 2011; Starkey et al., 2004).

2. Method

2.1. Research design

As mentioned above, the design of the current study comprised three conditions : a control group (CG), a first experimental group in which the games were implemented only at kindergarten (KG), and a second experimental group in which the games were implemented at kindergarten and provided to the families (KFG). Because of practical and ethical barriers in the countries where the study took place, a *controlled before and after quasi-experimental study* was conducted instead of a randomized controlled trial (Grimshaw, Campbell, Eccles, & Steen, 2000). In quasi-experimental designs – in which the participants are not randomly assigned to the experimental conditions – the estimate of effect cannot be attributed to the intervention with the same confidence as in randomized control trials. More specifically, a self-selection bias can occur. However, quasi-experimental designs with experimental and control groups that are pre- and post-tested can still provide information on the effectiveness of an intervention (Institute of Education Sciences, 2018, IES). Moreover, using multilevel analysis and introducing the mean performance of the classes as a level-2 variable (see below) allow partial control of the teacher effect that could arise from the self-selection bias (Chetty, Friedman, & Rockoff, 2014).

2.2. Participants

Kindergartens with mixed socioeconomic backgrounds were contacted in the local areas of the four participating universities[1] . Teachers interested in the study were asked to participate to one of the three conditions (CG, KG, and KFG) on a voluntary basis with the constraint there had to be at least two classes for each condition in each country. The study was approved by the Ethics Review Panel of the University that coordinated the project (University of Luxembourg). In every country, research permits were requested and received from the school authorities and from the children's parents.

A total of 23 kindergartens (3 in France, 11 in Luxembourg, 3 in Switzerland and 6 in Belgium), 46 teachers and 724 children initially participated in the study. There were from one to seven participating teachers per kindergarten. Table 1 shows the participation flowchart.

Before the pretest, teachers were asked to indicate which children had an attested important developmental disorder (trisomy/autism/severe developmental delay) or were new arrivals in the country and did not speak the language used at school. According to these criteria, 15 children were *a priori* excluded. Additionally, 140 more children were *a posteriori* excluded because of missing data for different reasons: 65 children were not present on the pretest day or the pretest was stopped because they did not seem to understand any questions, 75 children were not present on the posttest day or the posttest was stopped. Due to maternity leave, one teacher (and her 15 pupils) stopped participating during the implementation of the intervention (Table 1). Sample attrition rates were, respectively, 19.7% for the entire sample and 17.6%, 23.1%, and 18.3% for CG, KG, and KFG. No differential bias was observed in attrition across the CG, KG, and KFG conditions on gender ($F = 0.36$, ddl = 2, $p > .05$) and kindergarten year ($F = 2.40$, ddl = 2, $p > .05$) but was observed for age ($F = 3.14$, ddl = 2, $p < .05$). The pupils lost for the CG condition were slightly younger (mean = 56.9 months, SD = 12.3) than the pupils lost for the KG condition (mean = 61.6 months, SD = 7.5) and KFG condition (mean = 60.8 months, SD = 7.5).

Of the 569 remaining children, 238 (56.7% boys) were in prekindergarten and 331 (51.1% boys) were in kindergarten. The mean age at pretest was 55.5 months (SD = 4.7) for prekindergarten children and 65.8 months (SD = 4.9) for kindergarten children.

To consider important sociodemographic variables in the analyses, a survey was sent to the families of the children who participated in the experiment. Based on the 358 returned surveys for the children who had a pretest and a posttest score, 34.4% did not speak at home the language used at school. Socioeconomic and cultural status was calculated by averaging three standardized indicators: the social position index (Rocher, 2016) derived from parents' occupational status, the parental education level, and the home surface.[2] The mean socioeconomic and cultural status composite index was -0.03 (SD = 0.8). Sociodemographic data for each experimental condition is presented in the description of the analytic sample.

2.3. Early numerical skills assessment

Regarding the score to employ, the item response theory (IRT) was used in the current study. This paradigm has been proposed as an alternative to the classical theory (or true score theory), used in many previous studies measuring the effects of an intervention. The classical theory postulates that obtaining an error-free measure of the true competence of a subject is impossible. In practice, after having checked that the scale has an acceptable internal consistency (in general through the Cronbach's alpha coefficient), the correct answers that are given by the individuals are added without weighting and the observed score is associated with a more or less

---

[1] University of Liege (Belgium), university of Lorraine (France), university of Luxembourg (Luxembourg), University of Teacher Education from State of Vaud (Switzerland)

[2] Indexes of socioeconomic and cultural status have been proposed as alternative or complementary indexes relative to the socioprofessional category or salary indexes frequently used. Socioeconomic and cultural indexes provide more detailed and more precise information on the social environment in which the pupils evolve, and thus on their predisposition to academic success. In addition to professional or economic indications, socioeconomic and cultural status indexes also consider parents' educational level and housing conditions. Such an index is used, for example, in the PISA studies (index of economic, social and cultural status, ESCS).

Table 1
Participation flowchart of the study.

	CG	KG	KFG	Total
Initial state	13 classrooms 202 children	16 classrooms 244 children	17 classrooms 278 children	46 classrooms 724 children
Remaining pupils after *a priori* exclusion	13 classrooms 198 children (99.0%)	16 classrooms 238 children (97.5%)	17 classrooms 273 children (98.2%)	46 classrooms 709 children (97.9%)
Remaining pupils after lost by pretest	13 classrooms 180 children (90.0%)	16 classrooms 213 children (87.3%)	17 classrooms 251 children (90.3%)	46 classrooms 644 children (89.0%)
Remaining pupils after lost during intervention or posttest	13 classrooms 163 children (81.5%)	15 classrooms 183 children (75.0%)	17 classrooms 223 children (80.2%)	45 classrooms 569 children (78.6%)

important measurement error. Conversely, the IRT estimates the probability that a student will give a correct answer to a specific item, depending on the specific characteristics of the student and those specific to the item, leading to a weighted score (for more details, see Hambleton & Swaminathan, 1985).

In the current study, early numerical skills of pupils were measured by means of a test largely inspired by the TEDI-MATH (Van Nieuwenhoven, Grégoire, & Noël, 2001) and the TEMA-3 (Ginsburg & Baroody, 2003) batteries. TEDI-MATH enables the diagnosis of mathematical learning disabilities in children aged 5–8 years, and TEMA-3 provides a comprehensive assessment of mathematical skills for children aged 3–8 years. These two validated instruments were selected for their quality shown in the French-speaking (for the TEDI-MATH) and in the English-speaking (for the TEMA-3) contexts. They both allow the measurement of a large span of numerical skills without being over-aligned with intervention content. Some tasks were adapted to match the age of the target audience. Special attention was given to the intelligibility of the oral instructions, to avoid any misunderstanding of problems. The final version of the instrument (Table 2) used in this study comprised 34 items: 33 dichotomously coded items (0 = fail, 1 = pass) and 1 partial credit item (asking to count as far as possible: 0 = unable to count from1 to 10, 1 = able to count from 1 to 10, 2 = able to count from 1 to 20, 3 = able to count from 1 to 30 and 4 = able to count after 30). The reliability of the scale was good, with a Cronbach's alpha of 0.93. The same instrument was used at pretest and posttest.

An IRT model with partial credit including difficulty and discrimination parameters (Birnbaum, 1968) was used to estimate pupil competence and item parameters of the test. The analyses were conducted using Conquest 2.0 software, which generated a Warm's mean likelihood estimation (WLE) for each pupil. The difficulty and discrimination parameters of the items estimated during the pretest were anchored to compute the WLE scores for the posttest.

Regarding the assessment procedure, at pretest and posttest, children were individually tested in a quiet room/area of their school building by research assistants trained for the purpose. The test was administered in two sessions of approximately 15 min, depending on each child's attentional resources. The pretest took place just before the intervention (in Fall) and the posttest was conducted just after the intervention. In each condition, pretests and posttests were spread over 2 weeks.

2.4. Interventions

Intervention in the two experimental conditions (KG and KFG) comprised playing one mathematical game per week, for 8 weeks. These eight games are described in Table 3. They were selected for their international nature to facilitate their appropriation by the families, adapted by the research team to target specific numerical skills, and developed to be reproduced and distributed to the families at a low cost. For each game, both an easier and a more difficult version were prepared (Table 3) to help the teacher adapt the objectives according to the children's level. It was also specified that children with very little number knowledge could play most of the games with even smaller numbers than the numbers planned in the easier versions. To facilitate this, one large blank die was provided to each treatment class.

The teachers of both experimental groups participated in professional development sessions delivered by members of the research team. The first professional development workshop took place before the intervention and focused on early number skills acquisition, exploration of the first four games, and school–family relationships. For each game, the teachers received a summary sheet describing the rules, outlining the mathematical objectives to be pursued, and explaining how to adapt the games to make them easier or more difficult. The difficulties the children might encounter, the help that could be provided, and the learning to be formalized after each game were also presented in the sheets and discussed with the teachers. The same sheets were provided to the teachers of the treatment conditions in the four participating countries. During the first workshop, the sheets of the first four games were distributed to the teachers, who then discovered the games' stakes in workshop sessions, during which they simulated the activities and received feedback from the researchers. The other training sessions (one or two sessions depending on the country, for the same duration of 8 h for each teacher) took place during the middle of the intervention and began with the feedback of the teachers on the games that had been implemented up to that point. The researchers asked the teachers to describe how the games went and how the children managed. Next, the other four games were presented, and the games' stakes were again explored in workshop sessions.

During the intervention period, teachers of both experimental groups were asked to introduce each new game at the beginning of the week, then to make pupils play in small groups of two to four children while they walked around the groups. The teachers were also asked to formalize the learning targeted in each game, either collectively after the workshops or directly within the small groups when passing by a group ending a game. Teachers were asked to organize the learning in order to ensure that each child was involved four times, for 20 min each time, for each of the eight games. In both experimental groups, the games were implemented during school time, to a large extent instead of the usual math activities.

In the KFG group, each child took home the easiest or more difficult version – based on the teacher's evaluation – of the weekly game after it had been presented and played at kindergarten. A communication log accompanied each game to promote school-family exchange and to collect research data (e.g., playing time, difficulties encountered, pleasure experienced). Teachers were asked to make at least one comment per week concerning how the child appreciated or succeeded in playing the game. Parents were asked to report, through multiple-choice answers and pictograms, information concerning the pleasure and the ease their child exhibited, and the days their child had played.

In the CG, teachers were asked to continue teaching in the usual manner. In the countries where the study was conducted, the fundamental objectives regarding number and operations domains for

Table 2
Description of the early numerical skills assessment items.

Skills	Number of items	Description
Oral counting	7	Children are asked to count as far as possible (they are stopped at 42), up to/from a number (up to 6, up to 9, from 3, from 7, from 4 to 8, from 7 to 10).
Enumeration	7	Children have to enumerate linear and dispersed collections of elements (8 aligned rabbits, 6 dispersed sheeps); to give specific numbers of tokens to the experimenter (5, 8); to collect elements in order so there are as many as the farmers presented (5 hats, 7 rakes).
Conservation of numerical identity	3	Children are shown one row of 6 tokens that were previously enumerated and are asked how many there are after they are moved into a circle; children are shown two rows of 7 tokens facing each other and after the experimenter moved items of one row closer together, children have to say whether there is still the same number of tokens or not and to explain why.
Seriation	1	Children are asked to put cards with different quantities of elements from the smallest to the largest quantities (3 cards with 1, 3 and 6 frogs).
Numerical inclusion	2	A closed box with 6 rabbits inside is presented to the children who are asked if there are enough to take so many of them (8, 7).
Magnitude comparisons	4	Children are shown two collections of elements varying in size and differently dispersed and have to say on which side there are more (8 versus 7 dots similarly dispersed and of same size; 6 closely spaced versus 6 distant dots of same size; 7 closely spaced and smaller dots versus 6 distant and larger dots; 8 distant and smaller dots versus 8 closely spaced and larger dots).
Additions of concrete elements	3	Experimenter takes a small number of tokens in each hand, shows them to the children, hides both quantities in one hand and asks the children to say how many there are in all (2 and 1; 3 and 2; 5 and 3).
Story problems	3	Children are told addition and subtraction story problems such as “there were four hens in a courtyard; three more arrived; how many are there in total?” (4 and 3; 2 out of 6; 8 are 5 and . . .).
Number decomposition	4	Experimenter shows a certain number of rabbits to the children, hides a part of them while children close their eyes, after which children have to say how many are hidden by looking at the remaining elements (5 are 2 and. . .; 5 are 4 and. . .; 9 are 5 and. . .; 9 are 3 and. . .).

the end of kindergarten are, overall, to expose children to numbers up to approximately 10 through oral counting, enumeration, comparison, seriation, and estimation activities and to informally solve small additive and subtractive problems. The learning objectives are differentiated according to the pupils’ skills, with numerous pupils also employing larger numbers.

2.5. Fidelity of implementation at kindergarten and at home

Regarding the fidelity of implementation at kindergarten, all but one of the teachers[3] from the two experimental groups participated in the professional development sessions. They all received the aforementioned content and material. During the training sessions, the researchers stressed the importance of reliably implementing the intervention. In terms of dosage, teachers were asked to report weekly the number and duration of the game sessions. Because data was missing in many of these documents, teachers were asked afterward if they had been able to implement the intervention at the required exposure (4 × 20 min/week). Almost all the teachers reported that they did, and some of them even slightly more. Several teachers, however, reported that such a dosage would not be possible long term because of the kindergarten schedule and the other learning objectives. Since all but one of the teachers participated in all of the professional development sessions, and because the reported dosage of games was overall very close to the four-times-a-week for 20 min stipulation, all the treatment classes were considered as having completed the intervention with adequate implementation.

Regarding the KFG condition, the communication logs that accompanied each game provided at home were used as an indication of the home implementation of the games. Teachers were asked to collect the logs at the end of each week and return them to the research team. Not all logs were recovered because several parents did not return them and because some teachers either did not collect or misplaced them. Table 4 displays how many logs were collected from how many KFG children. From the 223 children in the KFG, at least one log was returned from 129 children (57.8%), while a majority of them returned eight, seven, or six logs. Based on the available logs, the number of times parents reported playing varied from 1 to 50 times over the 8 weeks (one time = one occurrence of one game), with a mean of 19.7 times (SD = 15.3). Besides the CG and KG groups, analyses were conducted considering two subsamples of KFG children: those 75 students (called KFG with proof) whose logs indicated that they had played at least 16 times during the intervention period (an average of twice a week) and the 148 others (called KFG no proof) who played less than 16 times, or for which no log was returned.

2.6. Analytic sample

The analytic sample included the 569 pupils who had pretest and posttest scores. As only 358 families (63%) returned the required sociodemographic information, family background information (i.e. socioeconomic and cultural status, language-mostly-spoken-at-home) were missing. To retain the full sample of 569 students, multiple imputation (Rubin, 1987) was done using the MICE (Multivariate Imputation by Chained Equations) algorithm in IBM SPSS 24. For the imputation model, we entered several parental variables that were available from the questionnaire (educational expectations, formal and informal literacy practices at home, formal and informal numeracy practices at home, involvement at school, self-efficacy concerning learning activities, parental role, confidence in the school, frequency of school-family communications). Five imputed files were produced and modeling results were automatically combined. Pooled descriptive statistics of the analytic sample are reported in Table 5.

As IBM SPSS 24 does surprisingly not provide pooled means difference testing parameters, statistical comparison across the three (CG/GS/KFG Total) or four (CG/GS/KFG with proof/KFG no proof) experimental conditions was made through regression using Mplus 8 software (Muthén & Muthén, 1998–2017) and dummy variables for group belonging. For the 3-groups comparison, statistically significant differences were observed for socioeconomic and cultural status between KFG and KG ($b = -0.272$, pMuthén = 0.004) and for language at home between CG and KG ($b = -0.131$, $p = 0.040$) and between CG and KFG ($b = -0.180$, $p = 0.007$). All other differences were not statistically significant at $p < 0.05$. Importantly, mean

---

[3] One teacher was hired after the 1st professional development day.

Table 3
Description of the eight math games and the targeted skills.

Games (and targeted skills)	Description and rules
1. *Ducks* (oral counting and enumeration)	The duck game consists of a simple linear board game with empty cells on which children move forward from left to right according to the number of dots obtained on a dice. There are bonus – throw the dice again – and malus – move one cell backward – cells on the path. The winner is the first player to arrive at the final cell. The easier version comprises 20 cells and is planned to be played with one dice up to six dots, whereas the more difficult version comprises 40 cells in order to play with two dice.
2. *Domino* (oral counting and enumeration, independence between the number and the area occupied by the elements)	As in a classical domino, each player receives seven dominoes pieces and the remaining pieces compose the deck. In turn, each player may place one piece if one of its sides contains the same number of items as one piece at the extremity of the domino path arranged on the table. Instead of regular dots patterns, this domino involves collections of different shapes varying in size and randomly displayed, in order to make enumeration necessary to check whether two collections have the same number of elements. Indeed, in a classical domino, matching might rely on a solely perceptive comparison. The winner is the first player to place all his pieces. The easier version involves quantities up to 6 and the more difficult version quantities up to 10.
3. *Battle* (oral counting, enumeration, independence between the number and the area occupied by the elements, magnitude comparison)	As in a classical battle, the cards are first shared between the players. Then, they simultaneously turn over the card on the top of their deck. Together, they determine who has the larger quantity and the one who has it takes all the cards. When two have the same largest quantity, there is a battle. As in the dominoes and for the same reasons, the collections of elements are different shapes varying in size and randomly displayed on the cards. The winner is the player with the most cards at the end. The easier version involves quantities up to 6 and the more difficult version quantities up to 12.
4. *The 11 game* (seriation, ordinal aspect of number, N + 1 and N-1 rule)	The eleven game involves four sets of cards going from 1 to 10 (easier version) or to 20 (more difficult version). Each set has a different color. One quarter of the cards is displayed as a deck in the center of the table and the rest is distributed across the players. The child who has the red 5 (or the red 11 in the more difficult version) puts it on the table. Then, in turn, children may play one or more cards by depositing the number that comes above or below the ones that are already deposited on the table. Players can also start a new set by playing another 5 (or 11). Each card indicates an Arabic number and the corresponding quantity of dots arranged in lines of maximum five. The winner is the first player to get rid of his cards.
5. *The Dragon's jail** (number composition and decomposition)	This game involves cardboard bricks, a dice with a dragon on the face representing 6 and a game board with a square of cells surrounding a dragon. On the square surrounding the dragon, a wall is drawn every five (more difficult version) or three (easier version) to ten cells. The cardboard bricks are designed to fill the cells and can be one to three (easier version) or to five (more difficult version) cells long, with a determined number of bricks of each length. Together, the players must confine the dragon in the jail by filling out the square with the bricks. In turn, each player rolls the dice (up to three dots in the easier version and up to five dots in the more difficult version). If the dice shows a number of dots, the player can take the corresponding length of bricks in one or more pieces and fill in the cells. The brick pieces may not cross the walls. If the dice falls on the dragon side, this one blows and remove from the square all the bricks that do not go from one wall to another. This incites children to fill in a set of cells one after another, which often requires decomposing the number of dots obtained on the dice in different brick configurations (for example, if three cells are left empty up to the next wall and a player obtains five on the dice, it will be in his best interest to take a brick of three and a brick of two, in order to fill the cells up to the wall). The players win together if they can confine the dragon, and the dragon wins if the children lack the bricks to lock him up.
6. *Rabbits and carrots* (number composition and decomposition)	This is a linear board game in which each player has two counters. On the linear path, there are empty cells as well as bonus (advance two or three cells) or malus (move backward two or three cells or back to the starting point) cells. In turn, each player rolls a dice and may decompose the obtained number to move their two counters separately, in order to avoid the malus cells and/or to reach the bonus cells. The easier version comprises 20 cells and could be played with a dice up to three dots, whereas the more difficult version comprised 40 cells and is planned to be played with a dice up to six dots. The winner is the first player to move his two counters to the final cell.
7. *Addition battle* (additions strategies)	This game is played as a classical battle game but instead of playing one card at a time, each player turns over the two cards at the top of their deck and adds them up. The player who has the higher total can take all the other cards. When two totals are both the largest, there is a battle. This game is played with real bridge cards from 1 to 3 (easier version) or to 6 or even more (more difficult version).
8. *The Extra-card* (number composition and decomposition)	This game is also played with real bridge cards. Depending on the number that is determined at the beginning of the game and which decompositions will be trained, all the cards below this number are included in the game (for example, when the number is five, the one, two, three and four of all four suits are kept). One "extra" card is also included, for example a queen of spades. All the cards are distributed across the players. Each player checks his cards to determine whether he has one or more pairs composing the number of interest and may lay it/them on the table. Then, in turn and as in the old maid game, each player draws a card from his neighbor. Each time one player can make a new pair composing the number of interest, he lays it on the table. The winner is the first player to get rid of his cards, while the player ending with the "extra card" loses. Training the decomposition of small numbers (three, four or five) makes up the easy version of this game whereas training the decomposition of larger numbers (from six to ten) makes up the more difficult version.

Note: *Adapted from an existing game (Van Lint & Defresne, 2020, retrieved from http://www.enseignement.be/index.php?page=26697).

Table 4
Frequency of children per number of logs returned over the 8 weeks.

	Number of logs returned per child over the 8 weeks								
	8	7	6	5	4	3	2	1	At least one log
Number of children per number of logs	41	28	12	11	9	2	11	15	129
Percentage of children per number of logs	31.8%	21.7%	9.3%	8.5%	7%	1.6%	8.5%	11.6%	57.8%

Table 5
Pooled descriptive statistics for the analytic sample.

	CG	KG	KFG			TOTAL (n = 569)
	Total(n = 163)	Total(n = 183)	Total(n = 223)	*With proof(n = 75)*	*No proof(n = 148)*	
Pre-K children (%)	44.8	43.2	38.6	*44.0*	*35.8*	41.8
Boys (%)	51.5	52.5	55.6	*54.7*	*56.1*	53.4
Mean age at pretest in months	61.1	61.4	61.8	*63.1*	*61.0*	61.5
Mean pretest WLE score	−0.13	0.08	0.07	*0.28*	*−0.04*	0.02
Pupils who mostly speak a foreign language at home (%)	23.1	36.2	41.1	*57.9*	*32.6*	34.3
Mean socioeconomic and cultural status index	−0.05	0.10	−0.17	*−0.21*	*−0.15*	−0.05

pretest scores for each experimental condition were not statistically different. This finding supports the baseline equivalence of the three groups. For the 4-groups comparison, statistically significant differences were observed for the pretest score between CG and KFG with proof (b = 0.403, p = 0.005) and between KFG no proof and KFG with proof (b = 0.317, p = 0.031). For the age variable, statistically significant differences were observed between CG and KFG with proof (b = 1.999, p = 0.040) and between KFG no proof and KFG with proof (b = 2.059, p = 0.037). For socioeconomic and cultural status, statistically significant differences were observed between KFG with proof and KG (b = -0.310, p = 0.011) and between KFG no proof and KG (b = -0.252, p = 0.018). For language at home, statistically significant differences were observed between CG and KG (b = -0.131, p = 0.040), between KFG with proof and KG (b = 0.217, p = 0.012), between KFG with proof and CG (b = 0.348, p = 0.000) and between KFG no proof and KFG with proof (b = 0.253, p = 0.021).

In other words, while the 3-groups comparison showed that KFG children reported a significantly lower socioeconomical and cultural status compared to KG children and that CG children were proportionally less likely than KG and KFG children to report speaking a foreign language at home, the splitting of the KFG group based on the logs data led to more heterogeneous groups. The 75 children for whom we know played at least 16 times at home (KFG with proof) were indeed different from the 148 children for whom we have no clue as to whether they played the games at home (KFG no proof). These 75 students were proportionally more likely to mostly speak a foreign language at home, but also initially more competent at the pretest.

2.7. Statistical analyses

In data sets with a hierarchical structure, individuals are in general not completely independent since people in the same unit tend to be similar to each other. According to Hox (2010), while standard statistical analyses rely on the assumption of the independence of observations, ignoring the hierarchical structure of the data leads to underestimating standard error and many spuriously significant results. To avoid aggregation bias and ecological fallacy, multilevel modeling was developed by several teams in the 1980s (Aitkin & Longford, 1986; Goldstein, 1986; Mason, Wong, & Entwisle, 1983; Raudenbusch & Bryk, 1986). Importantly, multilevel regression analyses are also designed to control for the initial differences that are statistically significant between the groups.

In the current study, multilevel regression equations were used to analyze the relationships between the intervention and the posttest scores. Six level-1 variables relating to the pupils were considered: pretest score, kindergarten year, age, gender, socioeconomic and cultural status, and language-mostly-spoken-at-home. Variables at level-1 were grand-mean centered. At level 2, three variables were included: the mean pretest score of the class and the experimental conditions coded as a dummy coding considering the KG as the reference group. This dummy coding allowed us to directly compare the KG and the CG condition as well as the potential added value of the KFG with proof and the KFG no proof compared with the KG.

Posttest scores were analyzed using the IBM SPSS 24 mixed syntax and a step-by-step procedure suggested by Hox (2010). Five multilevel models were built and compared. Model 1 is the *intercept-only model* (or unconditional model) with no explanatory variables. It shows how total variance of the posttest score is split into between-class variance (differences between classes) and within-class variance (differences between pupils). Model 2 includes all lower-level explanatory variables as fixed and with non-random slopes. In Model 3, the level-2 factors were added. This model examines whether the class-level explanatory variables explained between-group variation in the posttest score. Model 4 is the *random-coefficient model*. It assesses whether the slope of the pretest score variable had a significant variance component between the groups. In this model, the slope of the regression equation between the pretest score and the posttest score is allowed to vary from class to class. By introducing cross-interactions between the experimental condition and the pretest score, Model 5 allows to answer to the research hypotheses. It assesses whether the effects of the experimental conditions were the same for all children. In all models, the intercept refers to the expected mean posttest score when all predictors are zero.

3. Results

Model 1 (Table 6) shows how total variance of the posttest scores was distributed between intraclass variance and between-class variance. The intraclass correlation coefficient (ICC) of 0.21 (0.261/0.261 + 0.990) meant that one fifth of the posttest scores' variance was observed between classrooms. This relatively high ICC justifies the use of multilevel modeling. It suggests that aggregated (e.g., the mean class pretest score) or global (e.g., the treatment condition) variables at the level of the classroom could affect the posttest score of pupils.

Models 2–5 show how the intraclass and interclass variance proportions varied when explanatory variables were introduced. First, the shift from Model 1 to Model 2 led to the largest reductions in variance proportions. Model 2, which comprised math abil-

Table 6
Multilevel models explaining the posttest score (N = 569).

Model		1	2	3	4	5
Intercept		0.454**** (0.088)	0.653**** (0.034)	0.675**** (0.053)	0.665**** (0.051)	0.725**** (0.059)
Fixed Effects						
Pupil-level variables	Pretest score		0.925**** (0.027)	0.924**** (0.028)	0.939**** (0.037)	1.054**** (0.056)
	Kindergarten year		−0.052 (0.071)	−0.051 (0.071)	−0.057 (0.068)	−0.036 (0.067)
	Age		0.007 (0.005)	0.007 (0.005)	0.003 (0.005)	0.003 (0.005)
	Gender (girls as reference)		0.085* (0.045)	0.083* (0.045)	0.070 (0.044)	0.074* (0.043)
	Socioeconomic and cultural status		−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)
	Language-mostly-spoken-at-home		−0.001 (0.001)	−0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Class-level variables	Mean class pretest score			−0.015 (0.056)	0.021 (0.051)	0.036 (0.051)
	Dummy 1: CG (KG as reference)			−0.096 (0.073)	−0.084 (0.063)	−0.181** (0.081)
	Dummy 2: KFG with proof (KG as reference)			0.046 (0.088)	0.071 (0.081)	−0.007 (0.091)
	Dummy 3: KFG no proof (KG as reference)			−0.019 (0.074)	0.003 (0.065)	−0.082 (0.082)
	Pretest score*Dummy 1					−0.188** (0.078)
	Pretest score*Dummy 2					−0.160* (0.093)
	Pretest score*Dummy 3					−0.165** (0.082)
Random Effects						
Variance of the pupil-level residual errors		0.990**** (0.061)	0.280**** (0.017)	0.281**** (0.017)	0.263**** (0.017)	0.248**** (0.015)
Variance of the class-level residual errors		0.261*** (0.076)	0.016* (0.008)	0.012 (0.007)	0.009 (0.007)	0.016* (0.009)
Variance of the regression coefficients for pretest score across classes					0.015** (0.006)	0.017** (0.007)
Intercept-slope covariance					0.022 (0.013)	0.018* (0.009)
−2 log V		1673.15	915.01	912.28	889.37	882.16

Note: **** p < 0.001, *** p < 0.01, **p < 0.05, *p < 0.10.

ity at pretest, kindergarten year, age, gender, socioeconomic and cultural status, and language-mostly-spoken-at-home, explained 72% [(0.990−0.280)/0.990] of the within-class variance of posttest scores and 94% [(0.261−0.016)/0.261] of the between-class variance of posttest scores. In this model, only the pretest score variable was statistically significant at $p < 0.05$, suggesting that posttest scores were highly dependent on initial math ability. More precisely, when all other level-1 variables were kept constant, each increase of one unit of the pretest score was associated with an increase of 0.925 in the posttest score. Including class-level explanatory variables, Model 3 did not explain more within-class and between-class variances, and the effects of the experimental conditions were not statistically significant. Model 4 tested if the slope of the regression between pretest score and posttest score had a significant variance component between classrooms. This was partially the case as the variance of the regression coefficients for pretest scores across classes was statistically significant at $p < 0.05$ while the intercepts' slope covariance was not. In other words, there were some variations in the relation between initial and posttest math ability scores according to the class attended. In Model 5, which explained 75% [(0.990−0.248)/0.990] of the within-class variance of posttest scores and 94% [(0.261−0.016)/0.261] of the between-class variance of posttest scores, interaction terms between the pretest score and the treatment conditions were added. These interactions were negative and statistically significant at $p < 0.05$ for interactions with dummy 1 and dummy 3 (respectively, b = -0.188 and b = -0.165) and at $p < 0.010$ for the interaction pretest/dummy 2 (b = -0.160). Moreover, the negative parameter for dummy 1 (contrasting the CG and KG) was, this time, statistically significant at $p < 0.05$ (b = -0.181). These results mean that math gains were greater in the KG condition than in the control condition, but the negative parameter for the interaction between the pretest score and dummy 1 suggests that this positive effect differed according to the initial math ability of the pupils. The negative but non-significant parameter for dummy 2 and dummy 3 (contrasting the two KFG groups and KG) implies that the predicted added value of the family component of the intervention did not appear. Again, the negative parameters for the interaction between the pretest score and dummy 2 and the interaction between the pretest score and dummy 3 suggests that the positive effect of the math games varied according to the initial math ability of pupils.

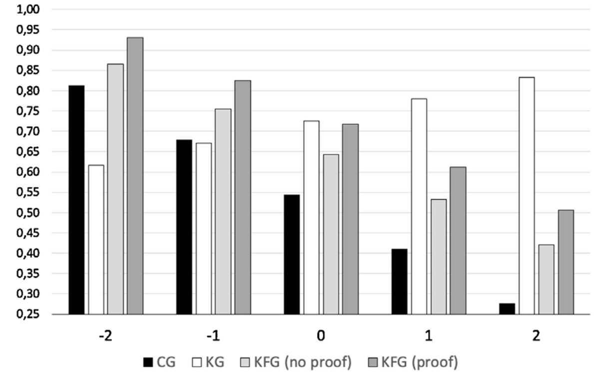

Fig. 1. Illustration of the fixed effects (mean gain scores) of Model 5 for arbitrary values of the pretest score (-2, -1, 0, 1, 2).

Fig. 1 helps visualize these fixed effects by using arbitrary and illustrative values of initial ability (-2=very low pretest score, -1=low pretest score, 0=average pretest score, 1=high pretest score, 2=very high pretest score). First, mean progress in the KG group was higher than in the other groups for children with average to very high ability at pretest but was weakest for children with low and very low initial ability. Second, mean progress in the two KFG conditions (with and without proof of play) was always higher than in the CG children. Third, the mean scores of the 75 KFG children for whom we are sure that they played the games at least 16 times were always higher than those for the 148 KFG children for whom we have no clue as to whether or not they played the games at home. Last but not least, while the initial math ability gaps between children tended to be reduced in the CG and KFG conditions, gaps between children tended to increase in the KG condition. In other words, while the intervention in the KG condition seems to be the most efficient in terms of overall mean gain, it increased the ability differences observed between children at pretest whereas they appeared to be narrowed by the KFG condition.

4. Discussion

4.1. Responses to hypotheses

As presented in the results section, multilevel Model 5 explained 75% of the within-class variance and 94% of the between-class variance of posttest scores. When the pretest score, kindergarten year, age, gender, socioeconomic and cultural status, language-mostly-spoken-at-home and mean score of the class at pretest were controlled, a mean negative net effect remained for the CG when the KG was the reference group, meaning that children who played numerical games at kindergarten progressed overall more than children in the control group. However, no statistical mean difference was found when contrasting the KFG with proof to the KG. This finding revealed that although children who played at kindergarten and who regularly played in their family also showed more gains than children in the CG, they did not progress significantly more, in mean, than children who played only at kindergarten. Thus, our first hypothesis was confirmed but not the second.

Nevertheless, significant interactions between the individual pretest score and the dummy variables indicated that gains in ability were not the same in the two experimental groups according to the children's initial level of performance. As illustrated in Fig. 1, when the games were implemented only at kindergarten, children who displayed average and above-average performance at pretest progressed more than in the CG. The opposite finding was observed for children with very low performance at pretest, who progressed less in the KG than in the CG. These findings suggest that a so-called *MatthewEffect* (Merton, 1968; Stanovich, 1986, 2000) occurred in the KG intervention, referring to the observation that *the rich tend to get richer and the poor poorer*. Our third hypothesis was thus invalidated in the KG condition.

Conversely, regarding the students of the KFG group who regularly played at home (twice per week on average over the intervention period), the gaps between children with lower and higher skills narrowed. Of particular interest, children who had low and very low number skills at the beginning of the study made more progress in the KFG group relative to their initial ability than their peers in the KG and CG groups. Notably, this interaction between the pretest score and regular play moments at home tended to be significant when controlling for all the level-1 and level-2 variables, including socioeconomic and cultural status and language-mostly-spoken-at-home. Thus, our third hypothesis was confirmed in this experimental condition. Children with low initial numerical skills particularly benefitted from playing the games at home, and this was the case above the socioeconomic, cultural, and linguistic background of the children.

4.2. Interpretation of results and pedagogical implications

4.2.1. kindergarten-based intervention

First, the significant mean net effect of the KG group is noteworthy. Indeed, this intervention was developed to be carried out by the classroom teacher in a class-wide setting and using simple materials. Such easily sustainable conditions allowed a majority of the children to have significantly greater gains than in the business-as-usual condition. More specifically, it should be highlighted that while the KG condition was more beneficial to initially average and above average children (mean gain between 0.73 and 0.83), it did still benefit very low achievers (mean gain of 0.62) whereas the CG condition was most beneficial for very low achievers (mean gain of 0.81) and least beneficial for very high achievers (mean gain of 0.28). This highlights that while the impact of the CG condition could at first sight be seen as more "democratic" or fair, the KG was actually more prone to "support all children to develop their full potential" (Cheminais, 2013, p.60). This observation also reminds us that in many business-as-usual conditions teachers tend to underestimate the mathematical competence of young children, and often fail to provide them with appropriate learning experiences (Engel, Claessens, Watts, & Farkas, 2016). It should also be noted that for experimental purposes, the intervention was implemented intensively during a short period of time (four 20-minutes sessions per week for 8 weeks). This imposed pacing allowed comparisons across the experimental groups, but many teachers reported that it was not easy to conduct within the kindergarten schedule and alongside the other learning objectives to be achieved. The encouraging results obtained in this experimental context give reasons to believe that the numerical games could perhaps also efficiently be used less intensively but over a longer period of time, or by varying between the different games. This could be more compatible with the teachers' schedules and better aligned with usual practices. In that perspective, providing all the professional development at the beginning of the school year would be more relevant rather than in waves.

Nevertheless, it should be understood why children with low initial ability benefitted less from the KG condition than from the business-as-usual condition. As indicated in the introduction, the interventions that have been implemented in prekindergarten and kindergarten have greatly differed in duration and in amount of targeted number skills. For instance, the lower progress of children with low initial ability in the KG condition is probably not due to a too high quantity of targeted numerical skills relative to the frequency and duration of the current intervention (approximately 640 min over 8 weeks). Indeed, Jordan, Glutting, Dyson, Hassinger-Das, and Irwin (2012) and Dyson et al. (2013) showed that at-risk kindergartners significantly benefitted from an intervention of a similar duration (720 min over 8 weeks) targeting even more numerical skills than the current intervention, such as the base-ten principle and problem solving. Instead, one explanation may be the pacing at which new games were introduced. As explained in the method section, the teachers were trained and received sheets containing advice to remedy several difficulties that struggling children could encounter in each game. However, the pacing of one game per week might not have made it easy for teachers to identify their pupils' performance levels and the amount of help required. Additionally, the rules and strategies of each game might not have been easy for children with low skills to understand, including some 4-year-old children who were probably not familiar with board games. Hence, pupils with low skills might have benefitted from the intervention implemented at a slower pace; they would have had more time to become comfortable with the game's stakes, and their teacher would have had better conditions to identify and overcome the barriers.

Another explanation might be that some of the proposed games were too difficult for children with low initial ability. For example, dragon's jail, rabbits and carrots, and the extra-card games not only involved the rather high-level ability of number decomposition but also other skills such as strategy, collaboration, or memory. Thus, low-performing children would possibly first need to concentrate on games that strictly target number knowledge. Another possibility is that even the easier versions of the implemented games (involving numbers up to 6) were still too difficult for some 4-year-old children who entered prekindergarten with no or very little number knowledge. For them, a more suitable approach could be to begin with games relying on subitizing, which can facilitate the meaningful learning of enumerating, number-comparison, and informal arithmetic skills (Baroody & Purpura, 2017; Clements, Sarama, & Macdonald, 2019). The span of numbers could then progressively be increased within the same game or as the children move from game to game.

Another possible explanation might be related to the instruction type and format of the intervention. With reference to the

Response to Intervention Model (Fuchs, Mock, Morgan, & Young, 2003), the KG intervention implemented in this study was a Tier 1 intervention: it addressed all the children in the classroom and was implemented with only the classroom teacher. Notably, several interventions have been found to be effective for young at-risk children (Mononen, Aunio, Koponen, & Aro, 2015; Nelson & McMaster, 2019). Some of them were Tier 1 interventions but were much longer than ours. For example, on the one hand, Clements and Sarama (2007) measured the effects of the *Building blocks* program, that targets number and geometry skills. It was implemented daily by classroom teachers among young at-risk children over a 25-week period. They observed that the disadvantaged children that composed the treatment group made significantly more progress than their peers in the comparison group. On the other hand, Clarke et al. (2011) also developed a comprehensive program (ELM) covering number and geometry skills provided daily for 25 weeks. These authors found that although this program was globally beneficial to at-risk students, it did not fully eliminate the gap between at-risk and average-achieving students. Consequently, they developed a Tier 2 kindergarten program (ROOTS, Clarke et al., 2016). This program exclusively focused on whole number understanding and was designed to be delivered by an instructional assistant in a small-group format, three times per week, for approximately 20 weeks. They observed that at-risk children progressed significantly more than those of the control group and that the gains of at-risk children in the treatment condition were greater than the gains of peers not at risk. Thus, this Tier 2 program resulted in an effective decrease of the achievement gap. Notably, many of the interventions that have been implemented and found to be effective for at-risk children were Tier 2 interventions, in which an additional adult worked with children individually (tutoring) or in small groups, often outside the classroom, on specific skills (Aunio & Mononen, 2017; Pellegrini, Lake, Inns, & Slavin, 2018; Shanley et al., 2017; Van Luit & Shopman, 2000). These interventions mainly used highly guided and explicit instruction, with frequent modeling, feedback, and revisions, together with regular monitoring of the children's progress (Mononen et al., 2015).

In the current KG intervention, children mainly played in small groups but with the teacher moving from group to group instead of his/her constant presence. Thus, the KG condition might not have benefitted low-performing children because the instruction was not sufficiently guided by the teacher. Knowing more precisely what the dominant instruction format was in the CG classrooms would be necessary to better understand why low-performing children in the KG condition progressed slightly less than their peers in the CG; however, some authors have suggested that working in small groups is generally rare in kindergarten classrooms (Wasik, 2008). Although some authors have claimed that working in that manner is a key component of high-quality instruction in early childhood (Hendrick & Weisman, 2006; Katz, 1995), other authors state that whole-group activities could better serve less-skilled children when considering the total amount of time that each setting requires (Shanahan, 2005; Sørensen & Hallinan, 1986). At least, working in small groups in a class-wide setting appears to be favorable only under certain conditions (Wasik, 2008). In the case of our intervention, teachers were free to manage the small groups as they wanted or were accustomed to. A posteriori, the complexity and importance of this dimension was probably underestimated in the current intervention. However, because having an additional adult for a small group of children several times a week is not always possible within the usual school resources, it would be very useful to better understand under which circumstances the numerical skills of low-performing children could be fostered by the classroom teacher within the classroom. Notable options could be for the teacher to conduct several short sessions per week of specific activities with the whole group, or to do so by staying with a small group of low-ability children while other pupils are engaged in activities that they can handle more autonomously.

4.2.2. kindergarten and home-based intervention

Regarding the KFG intervention, it is first important to point out that knowing precisely what occurs within the families during such interventions is not easy. In the current experiment, all the children of the KFG condition brought the weekly games and communication logs home and the parents were asked to report, through the multiple-choice questions and pictograms of the logs, information concerning the play moments at home. Of course, the returned logs might not accurately reflect reality. For example, teachers reported (in parallel interviews) that parents asked them questions about the games' rules without having filled out any logs, or that children told them they had played with their brothers and sisters, who did of course not fill out the logs. At the same time, some parents might have exaggerated the number of play moments because of social desirability. Despite these limitations, we chose to check the home-based implementation this way. In a study in which parents were asked if they agreed to participate in a condition in which their child was assigned based on the voluntary basis of the teacher, asking them to fill out the logs appeared less inquisitive than phone calls or home visits, as has sometimes been done. Asking the parents to report the required information through a phone application could have been an alternative that could have increased the information rate of the family implementation. Yet, it is interesting to see (Fig. 1) that the progress of the KFG without proof (played less than 16 times over the intervention period or no log returned), although less strong, was more similar to the progress of the KFG with proof (played at least 16 times) than to the two other conditions. Anyhow, for reasons of methodological credibility, the following comments focus on the children from the KFG whose returned logs attested to regular play moments at home (KFG with proof).

It should be specified that although the children of the KFG with proof had a higher mean pretest score, the multilevel regression analysis showed that the interaction between the pretest score and regular play moments at home was in favor of low and very low ability children and tended to be significant. Importantly, this effect appeared while controlling for pretest score, socioeconomical and cultural status and language-mostly-spoken-at-home. It is therefore noteworthy that the children of this group who had low and very low skills (see levels -2 and -1 on Fig. 1) made more progress than low and very low achievers in the other groups. In other words, regularly playing the games at home allowed for greater progress among low achievers than the two other conditions.

The greater progress that regular play moments at home furthered for low achievers might be explained by the higher exposure to the targeted skills and the individual scaffolding and feedback that family-based interventions allow (van Steensel, McElvany, Kurvers, & Herppich, 2011). These components are essential for enhancing the competencies of less skilled children (Berk & Winsler, 1995; Hattie & Timperley, 2007; Lembke, Hampton, & Beyers, 2012). We also postulate that giving the games to households' provided benchmarks to parents regarding the numerical concepts they could encourage and favored their involvement in their child's education (Cabus & Ariës, 2017; Cannon & Ginsburg, 2008).

Notably, the mean socioeconomical and cultural status of the KFG with proof was low (-0.21) and the group was composed of significantly more children speaking a foreign language at home. This reveals that the home-based intervention as was implemented here prompted parents who are not the closest to the school culture to get involved. By contrast, some studies where the home-based intervention group were the families who had responded to an invitation or participated in training sessions have found that the home-based group was socio-economically more advantaged than the other groups (Farrell, 2014; Jalbert & Pagani, 2007; Niklas et al., 2016; Sonnenschein et al., 2016). Recruiting more advantaged families not only raises ethical but also methodological questions, since the socioeconomic index is a moderator of the intervention's effects (Dulay et al., 2018; Manz, Hughes, Barnabas, Bracaliello, & Ginsburg-Block, 2010). In this study, parents from different backgrounds were able to support the learning of their child. This undermines the counters misgivings held by some authors concerned that home-based intervention programs tend to presuppose skills and knowledge that are not necessarily present "particularly in disadvantaged parents" and that cannot be fully developed simply through their participation in the intervention (van Steensel et al., 2011, p. 71).

Thus, just as it has been suggested that focusing on enjoyment rather than on educational value was more conducive to the recruitment of low-income families in home-based interventions (Lingwood et al., 2018), providing international games might be more attractive for the families who need it the most compared with a family mathematics curriculum (Starkey & Klein, 2000) or attending 12 evening classes (Hirsh et al., 2019). Of particular interest, in the literature regarding home-based literacy interventions, authors have emphasized that sensitive and emotional relationships between parents and their children should not be disrupted by the pressure that could arise from a teaching-learning situation (van Steensel et al., 2011). Resorting to games might perhaps better avoid this risk. Other materials have been used in previous studies, such as storybooks, Lego bricks or pretend kitchen sets (Chan, Praus-Singh, & Mazzocco, 2020; Gaylord, O'Rear, Hornburg, & McNeil, 2020). Knowing which material is better suited both to foster the children's number skills and to involve parents from different backgrounds would be valuable. For example, while some math storybooks have the potential to encourage abstract math reasoning, storybooks are not familiar to all cultures (Strasser & Lissi, 2009; Uscianowki, Almeda, & Ginsburg, 2020). In case of a play-based intervention, it should also be considered whether it would not be better to provide several different games over the intervention period rather than one single game. Notably, in Ramani and Scalise's study (2020), one game (a numerical magnitude comparison game) was provided for the 6-week intervention period, and no significant progress was observed in number knowledge compared to a control group, this despite a parental training session and weekly reminders. It could be that providing different numerical games across the intervention period, as done here, is more likely to incite regular math exchanges in families, which could be the main source of improvement in children's number skills. Hence, further studies, including qualitative studies, would be necessary to better understand the attractiveness of various materials to different families. Finally, even though methodological concerns lead researchers to circumscribe material and targeted skills, we should remember that supporting parent's early math engagement might be better achieved by incorporating math through a broad variety of materials and within parents' daily activities, talks and current roles (Mazzocco & Claessens, 2020).

Nevertheless, despite these encouraging results regarding low achievers, regular play moments at home did not result in all children in the group making significant progress. This absence of a mean added value from providing games at home relative to playing at kindergarten might be because children with average and above-average initial performance did not need, and therefore did not benefit from, this additional home-based training of basic number skills. It should be specified that in the KFG group with proof, the gaps between the high and low achievers narrowed not only because of the greater progress made by children with low initial ability, but also because of the lesser progress made by children with high initial ability (Fig. 1). This surprising observation might be explained by the fact that KFG condition required children who already had good numerical performance to play easy games at home. Comparatively, the middle and high achievers of the KG condition would have continued with their possibly more stimulating math family habits at home.

However, it is better to reach the goal of narrowing the gaps between children by increasing the performance of low achievers than by reducing the progress of high achievers. In this perspective, in a future implementation the numerical games could be provided only to the families of less-skilled children. This is in accordance with the principle of equity, which advocates giving more to those who need it most. Nevertheless, to prevent stigmatization of low-achieving children, children with average or above-average numerical skills could eventually still receive the games, betting that benefits could be achieved at other levels. These children could also be provided with more complex games, but with the risk of further widening the gap.

4.3. Synthesis, limits, and further prospects

In summary, the mean progress of KG and KFG children was significantly higher than CG children. More precisely, the KG intervention allowed children with average or above-average initial performance to make more progress in number skills than their peers in the CG. Conversely, the KG intervention as implemented here benefitted low achievers less than the business-as-usual condition. Further studies could investigate whether a numerical games program of this type implemented at a slower pace could allow these children to progress. However, more guided instruction by the teacher is probably a key factor to help these children catch up to their peers (Mononen et al., 2015). In that respect, providing clear landmarks to the teachers on the learning steps of the young pupils, as is the case in a learning-trajectories approach, would likely help teachers better identify the children in the class who need additional special attention and provide them with closer support (Clements & Sarama, 2009).

Notably, inquiring about the practices in the CG, which was not performed and constitutes a limitation of the study, would allow firmer conclusions to be drawn on the intervention's effects. To consider the teaching practices more objectively in a further study, a reliable observation tool such as the COEMET (Sarama & Clements, 2009) still must be developed or adapted to the local contexts in which this study was conducted. Such a tool would also provide a basis for teachers and schools for self-evaluation. In addition, checking the fidelity of implementation and studying what precisely occurred in the treatment classrooms through observations, audio recordings or teacher interviews would also have allowed more reliable and more in-depth information than the orally reported dosage by the teachers. Also, considering a 3-level model in the multilevel analyses would allow to control for potential kindergarten or country effects. However, in order to do so, a larger study should be conducted since there must be at least 20–30 units in each level of the model (Hox, 2010).

Furthermore, the findings regarding the family component compared with the kindergarten-based intervention were very instructive. Thus, testing the added value of a home-based component in addition to a mathematical kindergarten-based intervention, which has very rarely been conducted, is an experimental design to be further explored. Although the KFG condition did not allow children with middle and high performance to progress more than the kindergarten-based intervention, including the families by providing them numerical games allowed parents from different socioeconomic and linguistic backgrounds to get involved and was particularly beneficial to low achievers. In a further study, in order to provide the best help to all children, clearer benchmarks regarding the numbers and the version of the games that would be appropriate for each child according to their initial number skills would probably support the teachers and parents in making the most appropriate decisions. Such benchmarks would probably also reduce the differences that could arise during such an implementation between some teachers or parents who might be tempted to challenge their students/children and others who would prefer to consolidate the foundations. However, much remains unknown concerning the conditions for progress. Collecting multiple measures of what happens at home, including measures of interactions quality, would help build a more complete representation of the family support that is more prone to improve the children's number skills (Ramani & Scalise, 2020). For example, surveying whether providing games at home could initiate changes in parents' knowledge and practices would provide valuable additional information and help better understand the reasons why the intervention was effective for low achievers. Further research could also explore how to engage more parents and, more particularly, the parents of children of interest. Adapting other games that are closer to each family culture (e.g., Awalé, Bastra, Tablic) could be worthwhile. Finally, the maintenance of the benefits found here over the long term remains to be evaluated. Indeed, while studies indicate that involving parents in such programs produce long term effects (Ramey & Ramey, 1992; van Steensel et al., 2011), this has not always been found, for example with the Head Start program (Aughinbaugh, 2001).

It should also be remembered that such a study raises methodological and ethical concerns. For example, recruiting parents through an appeal or controlling the parental adhesion through attendance at training sessions or home visits creates a risk of truncating the sample and therefore the results. In some studies that have attempted to enhance the home numeracy/literacy environment in low-income families, special efforts have been made to reach the targeted audience using radio advertisements, door-to-door recruitment, an offer of meals and childcare during information sessions (Dulay et al., 2018; Hirsh et al., 2019). However, such implementations clearly require more than the usual school resources. Finally, in studies aiming to increase parental involvement, researchers should remember that parental involvement has a wide variety of expression. To conclude, although such home-based studies raise several questions, enhancing young children's numerical skills through a kindergarten-based intervention completed by a home play-based component appears to be a promising avenue for supporting each child's potential while reducing educational inequalities.

Declarations of interest

None.

Acknowledgments

This study was mainly supported by a grant from the University of Luxembourg (grant R-STR-3051-00-Z, 2015). The University of Teacher Education from State of Vaud, through the Lasalé Lab (Laboratoire sur l'accrochage scolaire et les alliances éducatives), supported the data collection in Switzerland. In Belgium, the data collection was supported by a grant delivered by the Sectorial Credits for Research in Human Sciences of the University of Liege.

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society Series A*, *149*, 1–43. http://dx.doi.org/10.2307/2981882

Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., . . . & von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Chilhood Research Quarterly*, *27*, 231–244.

Anderson, A. (1997). Families and mathematics: A study of parent-child interactions. *Journal for Research in Mathematics Education*, *28*, 484–511. http://dx.doi.org/10.2307/749684

Aughinbaugh, A. (2001). Does Head Start yield long-term benefits? *The Journal of Human Resources*, *36*(4), 641–665.

Aunio, P., & Mononen, R. (2017). The effects of educational computer game on low-performing children's early numeracy skills - an intervention study in a preschool setting. *European Journal of Special Needs Education*, *33*(5), 677–691. http://dx.doi.org/10.1080/08856257.2017.1412640

Baroody, A. J. (1987). *Children's mathematical thinking: A developmental framework for preschool, primary, and special education teachers*. New York, NY: Teachers College Press.

Baroody, A. J., & Purpura, D. J. (2017). Early number and operations: Whole numbers. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 308–354). Reston, VA: National Council of Teachers of Mathematics.

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction*, *30*(4), 435–470. http://dx.doi.org/10.1080/07370008.2012.720152

Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2018). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly*, *54*(2), 133–166. http://dx.doi.org/10.1002/rrq.229

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. Washington, DC: National Association for the Education of Young Children.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.

Bojorque, G., Torbeyns, J., Van Hoof, J., Van Nijlen, D., & Verschaffel, L. (2018). Effectiveness of the Building blocks program for enhancing Ecuadorian kindergartners' numerical competencies. *Early Childhood Research Quarterly*, *44*, 231–241. http://dx.doi.org/10.1016/j.ecresq.2017.12.009

Cabus, S. J., & Ariës, R. J. (2017). What do parents teach their children? The effects of parental involvement on student performance in Dutch compulsory education. *Educational Review*, *69*(3), 285–302. http://dx.doi.org/10.1080/00131911.2016.1208148

Cannon, J., & Ginsburg, H. (2008). "Doing the math": Maternal beliefs about early mathematics versus language learning. *Early Education and Development*, *19*(2), 238–260. http://dx.doi.org/10.1080/10409280801963913

Chan, J. Y. C., Praus-Singh, T., & Mazzocco, M. M. M. (2020). Parents' and young children's attention to mathematical features varies across play materials. *Early Childhood Research Quarterly*, *50*(P3), 65–77. http://dx.doi.org/10.1016/j.ecresq.2019.03.002

Cheminais, R. (2013). *Every child matters: A practical guide for teachers*. London and New York: Routledge.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *The American Economic Review*, *104*(9), 2593–2632. http://dx.doi.org/10.1257/aer.104.9.2593

Cheung, S. K., & McBride, C. (2017). Effectiveness of parent - child number board game playing in promoting Chinese kindergarteners'numeracy skills and mathematics interest. *Early Education and Development*, *28*, 572–589. http://dx.doi.org/10.1080/10409289.2016.1258932

Cheung, S. K., & McBride-Chang, C. (2015). Evaluation of a parent training program for promoting Filipino young children's number sense with number card games. *Child Studies in Asia-Pacific Contexts*, *5*, 39–49. http://dx.doi.org/10.5723/csac.2015.5.1.039

Clarke, B., Smolkowski, K., Baker, S., Fien, H., Doabler, C., & Chard, D. (2011). The impact of a comprehensive Tier I core kindergarten program on the achievement of students at risk in mathematics. *The Elementary School Journal*, *111*(4), 561–584. http://dx.doi.org/10.1086/659033

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Cary, M. S. (2016). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities*, *49*(2), 152–165. http://dx.doi.org/10.1177/0022219414538514

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building blocks project. *Journal for Research in Mathematics Education*, *38*(2), 136–163.

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Taylor & Francis.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, *333*, 968–970. http://dx.doi.org/10.1126/science.1204537

Clements, D. H., Sarama, J., & Macdonald, B. (2019). Subitizing: The neglected quantifier: Merging perspectives from psychology and mathematics education. In A. Norton, & M. Alibali (Eds.), *Constructing number* (pp. 13–45). Research in Mathematics Education. Springer, Cham. http://dx.doi.org/10.1007/978-3-030-00491-0_2

Clements, D. H., Sarama, J. H., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *42*, 127–166.

Codding, R. S., Chan-Iannetta, L., George, S., Ferreira, K., & Volpe, R. (2011). Early number skills: Examining the effects of class-wide interventions on kindergarten performance. *School Psychology Quarterly*, *26*(1), 85–96. http://dx.doi.org/10.1037/a0022661

Dulay, K. M., Cheung, S. K., Reyes, P., & McBride, C. (2018). Effects of parent coaching on filipino children's numeracy, language, and literacy skills. *Journal* *of Educational Psychology*,. Advance online publication. https://doi.org/10.1037/edu0000315.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. http://dx.doi.org/10.1037/0012-1649.43.6.1428

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*(2), 166–181. http://dx.doi.org/10.1177/0022219411410233

Engel, M., Claessens, A., Watts, T., & Farkas, G. (2016). Mathematics content coverage and student leanrning in kindergarten. *Educational Researcher*, *45*(5), 293–300.

Every Child a Chance Trust. (2009). *The long-term costs of numeracy difficulties* Retrieved from http://www.everychildachancetrust.org/counts/index.cfm.

Farrell, H. B. (2014). *A home literacy intervention to improve student reading and parental self-efficacy (Master's thesis). Available from ProQuest Dissertations and Theses database (UMI No: 1556003).*

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind Brain and Education*, *5*(3), 143–151. http://dx.doi.org/10.1111/j.1751-228X.2011.01122.x

Friso-van den Bos, I., Kroesbergen, E. H., & Van Luit, J. E. H. (2018). Counting and number line trainings in kindergarten: Effects on arithmetic performance and number sense. *Frontiers in Psychology*, *9*, 975. http://dx.doi.org/10.3389/fpsyg.2018.00975

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide (NCEE 2014-4005).* Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://whatworks.ed.gov.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice*, *18*, 157–171. http://dx.doi.org/10.1111/1540-5826.00072

Fuson, K. C., Richard, J., & Briars, D. J. (1982). The acquisition and elaboration of the number word sequence. In C. Brainerd (Ed.), *Progress in cognitive development. Children' s logical and mathematical cognition* (pp. 33–92). New York: Springer-Verlag.

Gaylord, O'Rear, C., Hornburg, C., & McNeil, N. (2020). Preferences for tactile and narrative counting books across parents with different education. *Early Childhood Research Quarterly*, *50*(P3), 29–39. http://dx.doi.org/10.1016/j.ecresq.2018.07.010

Gelmann, R., & Gallistel, C. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability, third edition (TEMA-3).* Austin, TX: Pro-Ed.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report - Giving Child and Youth Development Knowledge Away*, *22*, 1–24.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, *73*(1), 43–56. http://dx.doi.org/10.1093/biomet/73.1.43

Grimshaw, J., Campbell, M., Eccles, M., & Steen, N. (2000). Experimental designs for evaluating guideline implementation strategies. *Family Practice*, *17*, 11–16.

Hachey, A. C. (2013). Teachers' beliefs count: Teacher beliefs and practice in early childhood mathematics education. *Dialog*, *16*(3), 77–85.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. http://dx.doi.org/10.3102/003465430298487

Hendrick, J., & Weisman, P. (2006). *The whole child* (8th ed.). Upper Saddle River, NJ: Pearson Publishing.

Hill, N., & Craft, S. (2003). Parent–School involvement and school performance: Mediated pathways among socioeconomically comparable African American and Euro-American families. *Journal of Educational Psychology*, *95*(1), 74–83.

Hirsh, H. K., Richmond, M. K., Pampel, F. C., Jones, S. S., Molieri, A. C., & Jones, J. (2019). Results from a randomized controlled trial of the Motheread/Fatheread early literacy intervention: Evidence of impact in a rural community. *Early Education and Development*, *30*(2), 216–237. http://dx.doi.org/10.1080/10409289.2018.1544813

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. (2009). *A Mandate for playful learning in preschool: Applying the scientific evidence.* Oxford University Press.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3–34. http://dx.doi.org/10.1177/1529100615569721

Hox, J. J. (2010). *Multilevel analysis : Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Institute of Education Sciences. (2018). *What works clearinghouse* Retrieved from https://ies.ed.gov/ncee/wwc/.

Jalbert, J., & Pagani, L. (2007). La composante parentale d'un programme d'éveil numérique destiné aux enfants de milieu socioéconomiquement faible du niveau de la maternelle représente-t-elle une valeur ajoutée? *Revue Des Sciences de L'education*, *33*(1), 147–177. http://dx.doi.org/10.7202/016193ar

Jordan, N. C., Glutting, J., & Ramineni, C. (2009). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, *20*, 82–88. http://dx.doi.org/10.1016/j.lindif.2009.07.004

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2010). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*, 850–867. http://dx.doi.org/10.1037/a0014939

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology*, *104*(3), 647–660. http://dx.doi.org/10.1037/a0029018

Katz, L. (1995). *How can parents identify a high quality preschool program? ERIC Clearinghouse on Elementary and Early Childhood Education.*

Kleemans, T., Peeters, M., Segers, E., & Verhoeven, L. (2012). Child and home predictors of early numeracy skills in kindergarten. *Early Childhood Research Quarterly*, *27*, 471–477. http://dx.doi.org/10.1016/j.ecresq.2011.12.004

Klein, A., Starkey, P., Clements, D. H., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*(3), 155–178. http://dx.doi.org/10.1080/19345740802114533

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education*, *24*(2), 97–114.

Lee, J. S., & Ginsburg, H. P. (2009). Early childhood teachers' misconceptions about mathematics education for young children in the United States. *Australasian Journal of Early Childhood*, *34*(4), 37–45. http://dx.doi.org/10.1177/183693910903400406

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*, 1753–1767. http://dx.doi.org/10.1111/j.1467-8624.2010.01508.x

LeFevre, J., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science*, *41*, 55–66. http://dx.doi.org/10.1037/a0014532

Lembke, E., Hampton, D., & Beyers, S. J. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools*, *49*, 257–272. http://dx.doi.org/10.1002/pits.21596

Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology*, *46*, 1309–1319. http://dx.doi.org/10.1037/a0019671

Lewis Presser, A. L., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big math for little kids: The effectiveness of a preschool and kindergarten mathematics curriculum. *Early Education and Development*, *26*(3), 399–426. http://dx.doi.org/10.1080/10409289.2015.994451

Lingwood, J., Billington, J., & Rowland, C. (2018). *Evaluating the effectiveness of a 'real-world' shared reading intervention for preschool children and their families: A randomized controlled trial* Preprint available from https://www.researchgate.net/publication/326365012.

Manz, P. H., Hughes, C., Barnabas, E., Bracaliello, C., & Ginsburg-Block, M. (2010). A descriptive review and meta-analysis of family-based emergent literacy interventions: To what extent is the research applicable to low-income, ethnic-minority or linguistically-diverse young children? *Early Childhood Research Quarterly*, *25*(4), 409–431. http://dx.doi.org/10.1016/j.ecresq.2010.03.002

Mason, W. M., Wong, G. Y., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 72–103). San Francisco: Jossey-Bass.

Mazzocco, M. M. M., & Claessens, A. (2020). Introduction to the special issue: Parents supporting early mathematical thinking. *Early Childhood Research Quarterly*, *50*(P3), 1–3. http://dx.doi.org/10.1016/j.ecresq.2019.07.007

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*, 56–63. http://dx.doi.org/10.1126/science.159.3810.56

Mononen, R., Aunio, P., Koponen, T., & Aro, M. (2015). A review of early numeracy interventions for children at risk in mathematics. *International Journal of Early Childhood Special Education*, *6*, 25–54. http://dx.doi.org/10.20489/intjecse.14355

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.)). Los Angeles, CA: Muthén & Muthén.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence* Retrieved from http://www.nctm.org/standards/focalpoints.aspx?id=282.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education. http://dx.doi.org/10.3102/0013189X08329195

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity.* National Academies Press.

Nelson, G., & McMaster, K. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology*, *111*(6), 1001–1022.

Niklas, F., & Schneider, W. (2014). Casting the die before the die is cast: The importance of the home numeracy environment for preschool children. *European Journal of Psychology of Education*, *29*(3), 327–345. http://dx.doi.org/10.1007/s10212-013-0201-6

Niklas, F., Cohrssen, C., & Tayler, C. (2016). Parents supporting learning: A non-intensive intervention supporting literacy and numeracy in the home learning environment. *International Journal of Early Years Education*, *24*, 121–142. http://dx.doi.org/10.1080/09669760.2016.1155147

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers.

Journal of Experimental Child Psychology, *152*, 278–293. http://dx.doi.org/10.1016/j.jecp.2016.07.011

Payet, J. (2017). *Ecole et familles : Une approche sociologique*. Louvain-la-Neuve: De Boeck Education.

Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2018). *Effective programs in elementary mathematics: A best-evidence synthesis* Retrieved from http://www.bestevidence.org/word/elem_math_Oct_8_2018.pdf.

Platas, L. M. (2014). The Mathematical Development Beliefs Survey: Validity and reliability of a measure of preschool teachers' beliefs about the learning and teaching of early mathematics. *Journal of Early Childhood Research*, *13*(3), 295–310. http://dx.doi.org/10.1177/1476718X14523746

Ramani, G. B., & Scalise, N. R. (2020). It's more than just fun and games: Play-based mathematics activities for Head Start families. *Early Childhood Research Quarterly*, *50*(P3), 78–89. http://dx.doi.org/10.1016/j.ecresq.2018.07.011

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. http://dx.doi.org/10.1111/j.1467-8624.2007.01131.x

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*(3), 146–159. http://dx.doi.org/10.1016/j.appdev.2011.02.005

Ramani, G. B., Rowe, M. L., Eason, S. H., & Leech, K. A. (2015). Math talk during informal learning activities in Head Start families. *Cognitive Development*, *35*, 15–33. http://dx.doi.org/10.1016/j.cogdev.2014.11.002

Ramey, S. L., & Ramey, C. T. (1992). Early educational intervention with disadvantaged children – To what effect? *Applied and Preventive Psychology*, *1*, 131–140.

Raudenbusch, S., & Bryk, A. (1986). A hierarchical model for examining school effects. *Sociology of Education*, *59*(1), 1–17. http://dx.doi.org/10.2307/2112482

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*, 1301–1308. http://dx.doi.org/10.1177/0956797612466268

Rocher, T. (2016). Construction d'un indice de position sociale des eleves. *Education & formations*, *90*, 5–27.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Sarama, J., & Clements, D. H. (2009). *Manual for classroom observation (COEMET)-Version 3. Unpublished version.*

Scalise, N., Daubert, E., & Ramani, G. B. (2017). Narrowing the early mathematics gap: A play-based intervention to promote low-income preschoolers' number skills. *Journal of Numerical Cognition*, *3*(3), 559–581. http://dx.doi.org/10.5964/jnc.v3i3.72

Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior*, *55*, 223–229. http://dx.doi.org/10.1016/j.chb.2015.09.013

Senechal, M., & Young, L. (2008). The effect of family literacy interventions on children's acquisition of reading from kindergarten to grade 3: A meta-analytic review. *Review of Educational Research*, *78*, 880–907. http://dx.doi.org/10.3102/0034654308320319

Shanahan, T. (2005). *The national reading panel report: Practical advice for teachers*. Naperville, IL: Learning Point Associates.

Shanley, L., Clarke, B., Doabler, C. T., Kurtz-Nelson, E., & Fien, H. (2017). Early number skills gains and mathematics achievement: Intervening to establish successful early mathematics trajectories. *The Journal of Special Education*, *51*(3), 177–188. http://dx.doi.org/10.1177/0022466917720455

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*, 655–661. http://dx.doi.org/10.1111/j.1467-7687.2008.00714.x

Skwarchuk, S. (2009). How do parents support preschoolers' numeracy learning experiences at home? *Early Childhood Education Journal*, *37*, 189–197. http://dx.doi.org/10.1007/s10643-009-0340-1

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, *78*(3), 427–515, https://doi.org/10.3 102/00346543083 17473.

Sørensen, A. B., & Hallinan, M. T. (1986). Effects of ability grouping on growth in academic achievement. *American Educational Research Journal*, *23*(4), 519–542. http://dx.doi.org/10.2307/1163088

Sonnenschein, S., Metzger, S. R., Dowling, R., & Simons, C. L. (2016). Extending an effective classroom-based math board game intervention to preschoolers' homes. *The Journal of Applied Research on Children : Informing Policy for Children at Risk*, *7*(2), article 1.

Sood, S., & Jitendra, A. K. (2011). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities*, *46*(4), 328–346. http://dx.doi.org/10.1177/0022219411422380

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407. http://dx.doi.org/10.1598/RRQ.21.4.1

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford.

Starkey, P., & Klein, A. (2000). Fostering parental support for children's mathematical development: An intervention with Head Start families. *Early Education and Development*, *11*, 659–680. http://dx.doi.org/10.1207/s15566935eed1105_7

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19*, 99–120. http://dx.doi.org/10.1016/j.ecresq.2004.01.002

Strasser, K., & Lissi, M. R. (2009). Home and instruction effects on emergent literacy in a sample of Chilean kindergarten children. *Scientific Studies of Reading*, *13*(2), 175–204. http://dx.doi.org/10.1080/10888430902769525Strasser

Uscianowki, C., Almeda, M. V., & Ginsburg, H. P. (2020). Differences in the complexity of math and literacy questions parents pose during storybook reading. *Early Childhood Research Quarterly*, *50*(P3), 40–50. http://dx.doi.org/10.1016/j.ecresq.2018.07.003

Van Lint & Defresne, retrieved from http://www.enseignement.be/index.php?page=26697.

Van Luit, J. E. H., & Shopman, E. A. M. (2000). Improving early numeracy of young children with special educational needs. *Remedial and Special Education*, *21*(1), 27–40. http://dx.doi.org/10.1177/074193250002100105

Van Nieuwenhoven, C., Grégoire, J., & Noël, M.-P. (2001). *TEDI-MATH, Test diagnostique des apprentissages de base en mathématiques*. Paris: Editions du Centre de Psychologie Appliquée.

van Steensel, R., McElvany, N., Kurvers, J., & Herppich, S. (2011). How effective are family literacy programs? *Review of Educational Research*, *81*(1), 69–96. http://dx.doi.org/10.3102/0034654310388819

Vandermaas-Peeler, M., Ferretti, L., & Loving, S. (2012). Playing the ladybug game: Parent guidance of young children's numeracy activities. *Early Child Development and Care*, *182*, 1289–1307. http://dx.doi.org/10.1080/03004430.2011.609617

Vandermaas-Peeler, M., Nelson, J., Bumpass, C., & Sassine, B. (2009). Numeracy-related exchanges in joint storybook reading and play. *International Journal of Early Years Education*, *17*(1), 67–84. http://dx.doi.org/10.1080/09669760802699910

Wang, A. H., Firmender, J. M., Power, J. R., & Byrnes, J. P. (2016). Understanding the program effectiveness of early mathematics interventions for prekindergarten and kindergarten environments: A meta-analytic review. *Early Education and Development*, *27*(5), 692–713. http://dx.doi.org/10.1080/10409289.2016.1116343

Wasik, B. (2008). When fewer is more: Small groups in early childhood classrooms. *Early Childhood Education Journal*, *35*(6), 515–521. http://dx.doi.org/10.1007/s10643-008-0245-4

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

Zhou, X., Huang, J., Wang, Z., Wang, B., Zhao, Z., Yang, L., et al. (2006). Parent-child interaction and children's number learning. *Early Child Development and Care*, *176*, 763–775.