No child left behind, nor singled out: is it possible to combine adaptive instruction and inclusive pedagogy in early math software?

Agneta Gulz^{1,2} · Magnus Haake¹

Received: 10 November 2020 / Accepted: 7 July 2021 / Published online: 2 August 2021
© The Author(s) 2021

Abstract

The article addresses the challenge of combining adaptive and inclusive instruction in early math software, that is, to provide different kinds of support and challenges to different individuals in response to their different needs—yet avoid exposing children (whether far behind or far ahead) as being different. Arguments for adaption as well as inclusion are discussed, and an evaluative user study is conducted in which 42 3- to 6-year-old preschool children made use of a digital play-&-learn game for early math designed to combine adaptive instruction with inclusion during a period of 6 weeks. Data logging, performance measures, observations of children playing, and interviews with teachers are used to evaluate whether the adaptive and inclusive strategies worked out as intended. Results indicate that the goals of inclusion as well as the goals of adaptivity were met. A preliminary conclusion is that it is possible to combine adaptation and inclusion in early math software.

Keywords Educational software · Adaptive instruction · Inclusive pedagogy · Early math · Preschool

Introduction

Compared to non-digital instructional material, a main potential benefit of educational software is the possibility of dynamic adaptation to each individual so that tasks, challenges, and support follow each individual's particular learning curve.

There is, however, a potential drawback of treating individuals differently in response to their differences as learners, in that some individuals may then feel singled out. Consider a simple example from classroom activities with non-digital

✉ Agneta Gulz
agneta.gulz@lucs.lu.se

¹ Lund University, Lund University Cognitive Science (LUCS), Lund, Sweden

² Department of Computer Science, Linköping University, Linköping, Sweden

materials: While handing out exercises to the whole class the teacher selects adapted tasks for some students. She provides one student with special, simplified exercises, tells a second student “For you it is sufficient if you complete the first two pages”, and provides a third student with a set of special, more advanced exercises. To be exposed as different, as in these examples, is potentially detrimental for learners, both those who struggle and those who excel. When a student is openly treated differently from others, there is a risk that she feels marginalized (Florian and Spratt 2013; Klingner and Vaughn 1999).

This paper discusses the possibility to have your cake and eat it: respond to individual differences between learners *and* avoid exposure and potential marginalization.

It should be pointed out that exposure of differences is not automatically avoided by the implementation of adaptive strategies in software—rather the opposite. The choice of what is presented to a student is made by the adaptive software and is, in immediate terms, less public than a teacher placing a specific book on a student’s desk or telling a student to work with a special set of tasks. However, once students look around and talk to others about what they are doing, it is generally clearly visible who is ‘far ahead’ and who is ‘far behind’. Stars, scores, awards, and other features that represent difficulty levels or other kinds of progress are common standard in educational adaptive software.

Such *scoring* or *award* features have merits in that they may stimulate and motivate students. However, there are learning domains where their drawbacks can be larger than their gains, and where exposure of who is ahead or who is behind ought to be avoided, particularly at young ages. We argue that early math for preschool children—which is in focus in this article—is one such domain.

In the following, we discuss motives for adaptive instruction as well as for inclusive instruction, first in general terms and then specifically for the area of early math. A number of early math games that use adaptive instruction—but are not inclusive in the sense used in this paper—are reviewed. Thereafter, we present an evaluative study of 42 4- to 6-year-olds’ use of an early math game that proposes to combine adaptivity and inclusion. Behavioural data logs together with data from observations and interviews with preschool teachers and children are analysed in order to evaluate to what extent the play-&-learn game in effect reaches both pedagogical goals—being adaptive as well as inclusive.

Review of relevant literature and central notions

Adaptive instruction

In a broad sense adaptive instruction refers to “[...] all instructional forms that accommodate the needs and abilities of different learners” (Lee and Park 2008, p. 470). According to Aleven (2015) an adaptive system is one that “[...] adjusts the course of instruction in nuanced and effective ways based on learner differences [...]” (p. 11). The term ‘adaptive instruction’ has also come to specifically designate

digital technologies and techniques that in some way observe aspects of a learner's behaviour and adjust what is presented to the learner based on those observations.

Among the learner differences that can be *adapted to* are cognitive abilities, metacognitive skills, affective states, social skills, and learning styles. Instructional variables that, in turn, can be *adapted* are, for example, feedback, content sequencing, presentational formats, and scaffolding (Shute and Zapata-Rivera 2012). The present article focuses on adaptive instruction—specifically adaptive feedback, scaffolding, and level of difficulty—with respect to cognitive abilities and the learner's current level of knowledge and understanding.

The dynamic input for an adaptive system can be quite simple, such as calculations of task performance and progress over time according to simple schemas. Implicitly, however, the system may involve complex models based on research findings regarding common learning trajectories in the domain in question (Ginsburg et al. 2013). In addition, even when the explicit focus is to adapt to *cognitive* aspects, some *motivational* aspects are implicitly accounted for, since exposure to tasks that are 'neither too simple nor too difficult' affects motivation in at least two ways. First, such exposure increases the likelihood that the learner experiences progress and with that a sense of meaningfulness in a learning activity. Experiencing a sense of increased competence is the single most important factor for the desire to continue to engage in an activity (Bransford et al. 1999; Cleary and Zimmerman 2012); this is also in line with how self-determination theory (Deci and Ryan 1985) lifts mastering of tasks as central for motivation. Second, adaptive feedback supports a cyclic feedback loop that is known to increase motivation (Cleary and Zimmerman 2012; Peirce et al. 2008).

Adaptive instruction capitalizes on the Vygotskian theory (Vygotski 1978) that an individual's responsiveness to various forms of assistance or *scaffolding* can be indicative of her future performance. Notably, scaffolding can be provided by a peer, a teacher, or another adult—also by software. In Vygotskian theory, the difference between what someone is able to learn or accomplish on her own and what she can learn or accomplish with assistance is termed the *zone of proximal development* (Vygotski 1978). Importantly, this zone changes with the learner's development and progress. This, in turn, puts forth *dynamic difficulty*—an adequate degree of difficulty being dynamically generated on the basis of the learner's current skills—as a central concept (Sampayo-Vargas et al. 2013).

Inclusive pedagogy

Florian and Spratt (2013, p. 119) define inclusive pedagogy as “[...] an approach to teaching and learning that supports teachers to respond to individual differences between learners but avoids the marginalization that can occur when students are treated differently”. In other words, it means that all students in a group—with differing abilities, strengths, as well as weaknesses—can participate in a given activity without being exposed as lagging behind or being ahead of their peers.

Usually, the term ‘inclusive pedagogy’ refers to the inclusion of students with special needs in an integrated educational environment (Roos 2019). Yet, in this

article, it is used to refer to the inclusion of all children in a group with their differing abilities and needs, whether they have specific learning disabilities or not. Our focus is on whether *inclusiveness*, in this sense, can be combined with *adaptivity* in early math software.

Arguments for adaptivity and inclusion in early math instruction

Several longitudinal studies have shown that early numeracy skills predict later achievement in school (e.g. Griffin 2007; Lepula and Hannula-Sormunen 2019; Missall et al. 2012). Children who enter school with poor basic numeracy abilities have difficulty benefiting from math instruction (Jordan et al. 2009). However, most of them have no developmental disorder but are low performers whose difficulties primarily stem from external factors such as low socioeconomic status (SES) and low exposure and little practice (Aubrey et al. 2006; Denton and West 2002). Without intervention and support, these children are likely to remain low performers throughout school (Clements and Sarama 2008), but successful interventions that enhance achievement have been demonstrated—most notably for children from disadvantaged socioeconomic backgrounds (Bullough et al. 2014; Duncan et al. 2007; Jordan et al. 2009, 2012b; Mononen et al. 2014; Praet and Desoete 2014). In other words, early math interventions for children at risk can change their prospects of academic success.

A central argument for the implementation of *adaptive instruction* in early math software is that individual differences are large within groups of similar-aged preschoolers (Denton and West 2002). Thus, there cannot be any ‘one-size-fits-all’ interventions, but adaptive instruction is necessary if all children are to receive adequate support and challenge.

The main motive for combining adaptivity with *inclusive instruction* (Florian and Spratt 2013) is to avoid early establishments of low *self-efficacy beliefs* (Bandura 1997) in math, i.e. weak beliefs in one’s own ability to learn and perform in math. Even though the developmental path of self-efficacy in preschoolers is not well explored, it is known that already by the age of 7 many children have low self-efficacy beliefs in mathematics (McLeod 1992). Given this, it makes sense to counteract early establishments, in preschoolers, of ideas of one-self as less capable than others in (early) math, while in effect the considerable variability in early math skills in preschoolers largely originates from differences in exposure and training (Denton and West 2002; Jordan et al. 2009; Mononen et al. 2014; Praet and Desoete 2014). An aggravating matter is that self-efficacy beliefs are known to influence actual performance (Rattan et al. 2012), that is, self-fulfilling prophecies and vicious circles of low expectations and low performance risk to be established. Also in broader terms, there is much evidence for intimate relations between affective and cognitive factors in young children who are in the early stages of learning mathematics (Batchelor et al. 2019).

Review of adaptive early math games

There is a vast number of digital games that target early math, and some make use of simple forms of ‘levelling’ where a set of accomplished tasks on one level unlocks the next. It is far less common with more complex adaptivity of the kinds described in the section *Adaptive Instruction*. The few examples that, to our knowledge, exist are researcher based rather than commercial. We here describe some of the most well-known.

In the *Number Race* game (Wilson et al. 2006), the task is to advance in a race against the computer by choosing the larger of two non-symbolically or symbolically represented numbers. Adaptive instruction is implemented with a multidimensional algorithm that continually estimates the current skill level of the player and correspondingly adapts difficulty with respect to (i) the numerical distance between the two numbers to be compared, (ii) the response time before the computer picks the right answer, and (iii) a ratio between symbolic and non-symbolic representations.

Dots2Track (Butterworth and Laurillard 2010) is designed to support understanding of the relation between the numerosity in a dot pattern, the position on a number line, and the symbolic representation as a digit. For an incorrect answer, there is feedback on the meaning of the incorrect answer and its relation to the corresponding correct answer. Instruction is adaptive in that it selects the next task or trial according to the learner’s performance so far. New dot patterns are only introduced (by the teacher) when the current dot patterns have been mastered a specified number of times.

Whereas *Number Race* and *Dots2Track* are developed with a focus on children with specific mathematics learning difficulties and target a relatively small set of skills, *MathemAntics* (Ginsburg et al. 2011) is a comprehensive educational software package for preschool through third grade that focuses on numbers and operations. It is composed of a series of game-like mathematics activities and allows for individualization in that it evaluates student performance and adjusts activities accordingly as well as provides adapted individualized feedback to the learner.

The play-&-learn game *Critter Corral* (Blair 2013) likewise targets the broad population of preschoolers in general. The game aims to support the development of a flexible understanding of number for young children to ensure they are adequately prepared when they start school. *Critter Corral* supports integration of multiple number concepts and different ways of representing magnitudes in that number words are linked with different representations of magnitudes and with number symbols. Learners practice estimation, one-to-one correspondence, size and number comparison, and the concepts of more and less, and are also supported in the development of the foundation for addition and subtraction. The game is specifically designed not to simply signal ‘right or ‘wrong’ but to provide informative feedback, not the least so-called implication feedback, to the learner, who is then prompted to correct mistakes. Adaptive instruction is implemented via stepwise increasing or decreasing the scaffolding that is provided together within the unlocking of gradually more difficult sub-games.

Number Sense Game, NSG, (Vanbecelaere et al. 2021), which includes a comparison game and a number line estimation game, comes in both an adaptive and non-adaptive version. The difficulty levels that are varied concern range (1–4, 1–9, and 5–18), display duration, number of anchor points on the number line, and type of representations (non-symbolic, symbolic, and mixed notation). Depending on success, the same difficulty level is repeated or not, and strong players are allowed to skip levels.

Math Shelf (Schacter et al. 2016), finally, is inspired by Montessori mathematics materials and includes a number of games and puzzles. The application first teaches the quantities 1 to 5, focussing on subitizing, ordering quantities, one-to-one counting, and matching different quantity representations. The games practice the connecting of number names to symbols, matching numerals to quantities, ordering numerals, and counting. When children demonstrate mastery of the numerals and quantities 1 to 5, the same mathematical skills are practiced for the numerals and quantities 1 to 9 with a new set of games and activities. Adaptivity is implemented as follows: the third time a child is unsuccessful at a task, a cue is provided. If she/he does not succeed at this attempt either, the task will reappear in that child's future game play sequence. The goal is to ensure that each student receives additional practice until he or she is able to complete the activity without scaffolds.

Turning to the concept of inclusive instruction as used in this article, none of the games presented above caters for this. They all explicitly *visualize* progress in that they map *stronger or weaker early math performance* towards the fulfilment of game goals, for example scoring with stars, getting access to a novel character when reaching a certain level, or unlocking novel sub-games after progress.

Magical garden: an early math game designed to be inclusive as well as adaptive

Magical Garden (hence MG) is developed by Educational Technology Group, Lund University and Linköping University, Sweden (Gulz and Haake 2019; Haake et al. 2015; Husain et al. 2015). MG (like Critter Corral) builds on work by Griffin et al. (1994) and the main goal is to support an integration of multiple number concepts and ways of representing numbers.

On the first, most basic levels, MG involves neither symbolic nor iconic representations of numbers, beyond spoken number words. All actions are performed on virtual objects. After these initial levels follows a stepwise linking of number words with both magnitude and a variety of different number representations, starting with iconic representations (e.g. virtual fingers) and progressing via semi-symbolic (e.g. virtual dices) to fully symbolic (Arabic numbers) representations. A particular focus is put on concepts such as higher/lower, longer/shorter, too few/too many, and more/less, along with relations between them, and with the introduction of the basis for addition and subtraction. Among the mathematical concepts and operations practiced in the game are 'subitizing', 'counting', 'correspondence', 'cardinality', 'ordering', and 'relative magnitude'.

Fig. 1 MG screenshots showing: (left) the initial quite barren garden; (mid) the three characters Panders the Panda, Mille the Mouse, and Igis the Hedgehog; (right) the later, gradually thriving garden

Furthermore, MG makes use of the pedagogical principle of *learning-by-teaching*, with the child taking on the instructor's role, helping a digital tutee (or *teachable agent*) (Biswas et al. 2001) solve progressively more difficult tasks. The child is introduced to three characters—a mouse, panda, and hedgehog—whose garden is barren and in desperate need of watering. The child chooses one as her friend, whom she will help collect water drops to bring the garden back to life (see Fig. 1). For educational benefits of learning-by-teaching in digital games, see Chin et al. 2013.

MG comprises 60 scenarios, ordered by difficulty defined by number range (1–4, 1–6, or 1–9), representation (fingers, patterns of dots, stripes, patterns of dots of different formats, number symbols), and method (counting, stepwise addition and subtraction, addition and subtraction). All scenarios can be presented through several sub-games that feature distinct narratives: e.g. a near-sighted bumblebee needing help to find the right flower or a treasure hunter needing help to reach one of several caves in a cliff by attaching balloons to her basket. Regardless of sub-game, any given scenario is always repeated in three successive pedagogical modes (Fig. 2): after having been introduced to the task the child practices on her own (mode 1); then the child shows her friend how to do the task (mode 2), and, finally, the child supervises her friend (i.e. the panda, the hedgehog, or the mouse) who attempts the task (mode 3). Supervising involves accepting the friend's answer (when judged correct) by pressing the button with the happy smiley, or otherwise pressing the button with the unhappy smiley (and, in some cases, also providing the correct answer). A benefit with the three subsequent modes is that they enable practice by repetition while still providing variation.

The overall aim of MG is to offer meaningful tasks in the domain of early math with respect to challenges and support for all 4- to 6-year-olds in a preschool group—including the most and least advanced—and cater for an inclusive pedagogy. The challenge is to actually treat children differently depending on their particular need, but not make the differential treatments explicitly visible. In this, MG is to our knowledge the only educational software that attempts to combine adaptive instruction with an explicit catering for inclusion.

Fig. 2 Screenshots from MG showing the three subsequent pedagogical modes reflecting the teachable agent paradigm: (*upper left*) overview of the sub-game Bird Hero where the task is to help the baby birds up in the tree to their parents; (*upper right*) child practices (*mode 1*); (*lower left*) child shows friend (Panders), who watches (*mode 2*); (*lower right*) child supervises friend (Panders), who proposes a button to press while the child either says no to this by pressing the button with the unhappy smiley or says yes by pressing the button with the happy smiley (*mode 3*)

Design for adaptivity in the game magical garden

MG is implemented with a state machine that keeps track of the progress of the child. Adaptivity manifests itself in that different individuals are steered through the game by different paths. The amount of repetition, and the places where repetition takes place, will differ, as well as the amount and kind of support at various levels. After each completion of one of the 60 scenarios, the following is evaluated: the degree of success with the just completed scenario, a weighted historical success rate in previous sessions including how many times the player has completed the scenario and in what ways, and how fast the child has advanced through the game. The result of the evaluation is used to decide on one of five options for which scenario will be presented next for the player: (i) repeating, i.e. practicing the same scenario; (ii) going back a ‘half’ level in difficulty, i.e. repeating the same scenario but with more scaffolding; (iii) going back an entire level in difficulty so that the child can practice and prepare for a new attempt on the higher level; (iv) progressing to the next difficulty level.

In this way some children get a substantial amount of training with a certain kind of task, whereas others quickly leave the same task behind. The basic logic is that a child who masters a pedagogical scenario moves to the next, while a child that has trouble repeats the same scenario (with different sub-games) until her performance improves. From the child's perspective, the mix of the sub-games played present variety, even if the pedagogical challenge is actually the same. MG relies on dynamic assessment, i.e. integrating assessment and instruction (Vygotski 1978), to address a 'potential for learning' rather than a 'static level of achievement'.

Design for inclusion in the game Magical Garden

Design for inclusion, as the term is used in this text, requires that instruction accommodates to learners' different needs and abilities, but also that learners are not exposed as being different from their peers—whether behind or ahead. MG attempts to cater for inclusion by making comparisons between children's early math competencies difficult: (i) The reward system (collecting water drops) is the same through all levels of play, with no additional rewards at higher levels. (ii) All successfully accomplished tasks yield the same number of water drops, independent of difficulty level; the number of collected drops reflects only the number of sub-games played successfully. (iii) The plants in the garden are randomly generated, so that all gardens look different regardless of how far one has proceeded. (iv) The different sub-games are distributed randomly over the 60 scenarios, which makes typical 'game level' comparisons difficult.

Research questions

A proof of concept of whether an educational game works as intended does, of course, not reside in design intentions, but in students' actual use, their learning experiences, and their teachers' views. Therefore, an empirical study was set up to investigate whether the design intentions regarding adaptivity and inclusion for Magical Garden were met. Would the combination of adaptivity and inclusion have the intended effects?

With respect to *adaptivity* we addressed the following three research questions:

RQ.1 Would all participating children stay motivated to use MG during the intervention?

RQ.2 Would all children be adequately scaffolded as well as challenged in the sense of neither getting stuck nor hitting the roof?

RQ.3 Would all children make progress with respect to early math skills?

With respect to *inclusiveness*, two further research questions were posed:

RQ.4 Would (some) children find themselves exposed as being behind or ahead in the game?

RQ.5 Would (some) children talk about game play experiences in ways that indicated that they compared themselves with regard to the early math skills?

Methods

Participants

Forty-two children (20 girls, 22 boys) from four different preschools in Sweden participated in the study. The age span was 3;9–6;5, with a peak around 5 years of age. All children spoke Swedish, but three of them did not have Swedish as their first language and another two children had limited linguistic abilities. The children came from socioeconomic and sociocultural low middle-class and high middle-class families.

Twenty-one of the children (10 girls, 11 boys), ages 4;2–6;2, completed pre- and post-tests targeting early math with an adapted subset of the *Number Sense Screener* (Jordan et al. 2012a). These were all children at two of the preschools. Due to logistic constraints, it was not possible to complete pre- and post-tests at the two other preschools. In line with the previous research (e.g. Griffin et al. 1994; Praet and Desoete 2014), the early math skills as measured by these tests displayed a large variability, with the scores on the pre-tests ranging from 4 to 17 (max score = 22). The variability was large also for children of similar age.

Procedure

The intervention with the play-&-learn game lasted for six consecutive weeks during which the children used MG during 20-min sessions three times a week. Each session took place in a separate room—commonly used for small-group activities—and involved a smaller group of three to four children with one iPad each, and one teacher or researcher. Depending on children's wishes they could use headsets. Researchers were present at the first session, at two sessions during the third and fifth week and at one session during the sixth week. During these occasions, the researchers made field notes while they observed children play, talked to children in pairs or small groups, and interviewed teachers.

Data collection methods

Game logs

The game logged interactions in the form of correct or incorrect answers on all tasks provided to the children and their digital friends, together with timestamps. For the analyses, these data were aggregated into a progression curve for each individual, that revealed whether they got stuck at certain levels (scenarios), if there were challenges (levels) left or not, and whether they were playing and making progress all through the 6 weeks.

For the log data to be accurate and reliable, it was crucial that the child logged in as a player was indeed the one actually playing, and that the children would not make moves for someone else on their tablets. They were not forbidden to talk and

discuss with each other during game play, or to show each other what happened on their screens—but were not allowed to act on another child's tablet. The preschool teachers were informed about the importance of this and were instructed to supervise it. One teacher expressed a concern at the start of the study “This will probably be very difficult for the children; they are used to play together on an iPad”—but in effect it turned out that the children accepted the rules very readily. Preschool teachers reported only two instances of having logged in a child on another child's account—discovered immediately by the children, one seeing that the name was wrong, the other not recognizing her garden. Otherwise, children, according to the teachers, had found it easy to follow the rule of only working on their own tablet. This was confirmed by researcher observations. Children engaged in small talk and discussions around their game play and what happened—but never interfered with another child's tablet. Taken together the teachers' reports and the researchers' observing and talking to children provided strong support that the events registered in the game logs represented the individual children's own actions.

External pre- and post-test

A subset of the standardized and widely used Number Sense Screener (Jordan et al. 2012a) provided a complementary measure of progress with respect to early number sense in relation to the in-game measure. The tasks in the pre- and post-tests included recognition of small quantities; counting items in sets; knowing that the final count word indicates how many are in the set (cardinality); discriminating between and comparing small quantities (which is more and less); mental operation on sets where items are added or taken away. For all tasks, instructions were orally given according to a script and supported by visual material. The maximum score was 22, and the scoring was guided by the Number Sense Screener (NSS) manual.

Researchers' observations of children playing

A main observational focus was the children's understanding of the overall narrative of the game and the different scenarios. Another focus was to what extent they exhibited motivation to engage in the game activities. We listened carefully to their conversations among one another and with the characters in the game, being attentive to if they talked in ways that indicated that they compared their early math skills, if they expressed that they were not capable to solve the game tasks, and if they indicated that they found themselves exposed as being behind or ahead in the game. Field notes were made during the observations.

Interviews with children

Researcher conducted interviews with all children except the child who quit after one session, in groups of 2–3 children at a time. Two central topics in these interviews were “How will you know if you are doing well in the game?” and “Can you know that you have come far in the game, and in that case how can you know this?”.

Interviews with preschool teachers

At each of the four preschools one of the researchers carried out interviews with the teachers. As noted above, the teachers had been instructed to be observant of (i) behavioural signs or conversations among children regarding comparisons of who has advanced most, (ii) children's talking or behaving in terms of competition, and (iii) children expressing low confidence in their ability to play MG.

In total, twelve teachers participated in the interviews, four from one preschool, two from another, and three each at the two remaining preschools. All interviews took place towards the end or just after the intervention period. In response to practical constraints, the interviews were sometimes conducted with a pair of teachers and sometimes with a single teacher. The interviews took the form of a conversation, guided by a set of questions that were posed in all interviews. See Appendix.

Results

Here, we present results as they relate to each of the five research questions.

Results on adaption (RQ.1–RQ.3)

RQ.1. Would all participating children stay motivated to use MG during the intervention?

According to the teachers as well as researchers' observations, the majority of children were pleased with playing the game. They were generally glad when it was their turn to have a play session, and overall children stayed very focussed on the activity. An observation made by teachers at two of the four preschools was that a couple of children at first found it difficult to calm down and focus. However, with the support of the teachers, in the course of one or two sessions they learned to do this, and subsequently seemed to enjoy the game sessions as an opportunity to relax and focus. In this context, it can be mentioned that the overall pace of MG—animations, music, ways of speaking—is somewhat slower than the majority of apps that children encounter.

Some of the younger children would talk extensively to the characters in the game, in particular their digital friend, and were clearly engaged in the narrative and what was happening. On some occasions it was difficult to convince some children that they had to quit playing. On the other hand, some of the older children were less enthusiastic. Two children (6-year-olds) told us that they found the game pretty boring because it was too simple, which was also indicated by the steep learning curves in their MG-logs.

RQ.2. Would all children be adequately scaffolded as well as challenged in the sense of neither getting stuck nor hitting the roof?

All children started with the same pedagogical scenario in the form of a very basic kind of task involving no representations but a set of four visual objects (balloons),

Fig. 3 Examples of different adaptivity patterns plotted as achieved level vs. number of rounds played. *A* (upper left) illustrates a child repeating scenario 13 (with different sub-games) in total five times; *B* (upper right) illustrates a child rapidly moving from one scenario to the next; *C* (lower left) illustrates one child that never ‘took off’; *D* (lower right) illustrates a variety of different adaptivity patterns

of which one, two, three, or four should be moved according to the task at hand. After this shared starting point, the log data show a large variety of different paths for different children (Fig. 3). Whereas some repeat a particular scenario several times, though with different sub-games (Fig. 3a), others move on quickly from one scenario to the next (Fig. 3b). It could also be inferred from the logs that all children, but one, made progress (Fig. 3c) and that the individual progress curves differed considerably (Fig. 3d). Five individual children (four 6-year-olds and one 5-year-old) stood out in that their progress curves were markedly faster through the levels as exemplified in Fig. 3b.

Looking at the group as a total, it is clear that all children, but one, made progress through the levels (pedagogical scenarios), but that the slopes of the curves varied considerably (cf. Figure 3d). Apart from the one child, who never ‘took off’ and thereafter quit playing (cf. Figure 3c), there are no individuals that got stuck never to go on and then quit. Conversely, there are no individuals for which the logs reveal that they ‘hit the roof’. Another indication that children were supported by the game to improve their skills was that the completion time for mode 2 was

shorter than for mode 1. The average completion time for mode 1, where the child acts alone, was 41 s. For mode 2—identical to mode 1 except that the digital friend watches the child acting—the average completion time dropped to 32 s.

RQ.3. Would all children make progress with respect to number sense and early math skills?

According to game logs, all children but one made progress through the levels in the game even though the pace of progress differed. There were also complementary results from an external pre- and post-test for 21 children (there were no differences between these 21 and the other half in terms of game progress,¹ age or gender). These results indicate improvement in early math skills for all 21 as measured by the subset of tasks from the Number Sense Screener (NSS). The overall improvement between the pre- and post-test (pre: $M=9.2$, $SD=3.63$; post: $M=13.2$, $SD=3.96$) showed a strong significant effect as evaluated by a paired t test ($t(20)=4.28$, $p<0.001$, Cohen's $d=0.93$). It should be emphasized that this cannot reliably be used as a *stand-alone* result because of obvious limitations in terms of a small number of participants and a lack of control group. It is nevertheless valuable as a complement to the game-data showing progress in the area for 41 of 42 children. It can also be noted that the time of the intervention was only 6 weeks which makes it unlikely that the progress between pre- and post-tests was only or primarily due to the children's general development.

Summary on adaptation (RQ.1–RQ.3)

Taken together, the log analyses and the observations by teachers and researchers indicate that the goal of presenting meaningful activities to an entire group of 4- to 6-year-olds in a Swedish preschool context was met. Except for one participant, all children, with their different age, gender, linguistic background, and previous knowledge were willing to use MG and did stay motivated to do so during the 6 weeks. All children seemed to be adequately scaffolded as well as challenged in the sense of neither getting stuck nor hitting the roof, and all children make progress in the game.

¹ Except that the child who did not make progress in the game at the start, and quit playing after the first game session, was among the 21 who did not complete the external pre- and post-tests.

Fig. 4 Number of water drops received after efficient task-solving time in Magical Garden for 15 min (left) and 30 min (right). The two timestamps represent the time spent on task solving in the three subsequent 'modes', i.e. time spent in the 'garden' is not included in these two time-measures

Results on inclusion (RQ.4–RQ.5)

RQ.4. Would (some) children find themselves exposed as being behind or ahead in the game? & RQ.5. Would (some) children talk about game play experiences in ways that indicated that they compared themselves with regard to early math skills?

To address the two research questions concerning inclusiveness, we relied on observations of children playing and talking to each other and interviews with the pre-school teachers, complemented with aspects of the log data. Since the answers to the two questions are often interwoven in each other in, for example, the same interview, we choose to present them together.

Log data

One central design feature of Magical Garden intended to support inclusion is that children gain water drops to use for their garden regardless of which pedagogical scenario they are working with and regardless of the pace of their progress. After each completed round the child and her digital friend receive three water drops. A round is completed when the child has produced three correct answers in each of the three modes. This happens regardless of whether the child has made many mistakes during the round or answered mostly or completely correctly on all tasks. Children who make few incorrect answers will advance *quicker through the scenario levels* than children who make many mistakes. But children who make many mistakes will receive *more water drops per scenario level* than children who make few mistakes and advance quicker through the scenarios. That is, the number of water drops collected, and the growth of the gardens, will not map 'skill'. Figure 4 illustrates this from a log perspective by plotting the number of water drops gained after a total of 15- and 30-min efficient task solving, in

relation to what we call in-game performance—the percentage of correct answers with regard to the total number of given answers. The left picture shows a fairly even distribution of water drops with regard to children's in-game performance after 15 min. As for the 30-min check point, shown to the right, the children with higher in-game performance have on average collected a larger number of water drops but there is no straight-on mapping between progress curve and number of water drops gained. The four children with the slowest progression have received an equal amount or even more water drops than some children with clearly faster progression.

Researcher observations of children's playing and small-group interviews with children

Did children discern the differences among themselves regarding progress rates and task difficulties with respect to math? Did game play invite them to compete and compare themselves in terms of early math skills? The researcher observations made clear that the children *did make comparisons* of several kinds. They often talked about the different looks of the gardens, and about the plants that grew there, for example: “Look I have two big candy canes in my garden, what about you, you have one, but only one”. However, such kinds of differences are not associated with learning progress. It may well be the slower or weaker child who has more candy canes in her garden. Since none of the different aspects that children compare are systematically related to relative progress (quicker, higher levels), there is little risk that any child, whether low-achiever or high-achiever, would self-label themselves as weak on the basis of these aspects.

Some children—primarily older ones—had intense discussions about what is required in order to enter a ‘new level’. As anticipated (and desired) their concept of a ‘new level’ did not refer to a more challenging pedagogical scenario, but to ‘a change of sub-game’. In effect, nothing particular is required to have a change of sub-game because this is a random feature in the game. But the discussions among children about this were engaged and the suggestions creative, for instance: “if one does not use up the water-drops in the watering can, there will not be a novel level”; “if Panders [the digital friend] does not learn well, there will not be a novel level”, etc.

One question discussed in the small-group interviews was “How will you know if you are doing well in the game?” Typical answers were “One gets to another level”; “There are novel things to try out”; “There are many flowers in the garden”. These answers indicate that the goal of inclusion is met. For the children ‘another level’ does not equal ‘another pedagogical scenario’ but ‘another sub-game’—something all children experience regardless of how well they do. Also, the number of flowers does not translate directly to speed of progress. Children were also asked if and how one can know that one has come far in the game. Again, answers varied, but these were common: “If you play for a long time, you will get far” (6 children); “If you get to novel levels” (5 children); “If you can water many things” (7 children); “If you play quickly” (3 children); “If you can do all levels” (3 children).

Researcher interviews with teachers

Turning to interviews with teachers, their answers confirmed the researchers' observations of the kinds of comparisons made by children. There was, indeed, teachers reported, quite a deal of 'comparison talk', in particular amongst older children, but few signs of competition.

The teachers agreed on the conclusion that in the game sessions with Magical Garden every child could be engaged and included. In effect, this applied even to two children who did participate in the gameplay sessions contrary to the teachers' expectations beforehand. Both children had communicative problems and one of them was diagnosed with autism. According to the teachers it had not happened before that these children had stayed focussed this long when using a game at the tablet. The child with autism was also seldom involved in the same activities as other children. In sum, teachers responded positively to how inclusive pedagogy was implemented and how well it worked.

Summary on inclusion (RQ.4–RQ.5)

Our preliminary conclusion is that the game does meet the goals of inclusion. Notably (but not surprisingly), children spontaneously found many ways to compare what happens on their respective screens, but none of these comparisons mapped to actual progress in the game with respect to early math. Furthermore, all children seemed to experience progress for themselves.

Discussion

General conclusion

Overall, the results from the study indicate that the design of the early math game Magical Garden, that strives to combine adaptivity and inclusion, had the intended effects on the participating children. The age span in the group was between 3;9 and 6;5 and the early math knowledge and skills at the start of the intervention varied considerably in the group. Even so, all children showed a willingness to use the game. The paths taken through the game depended on their learning curves and differed largely among children. Yet, all but one child made progress in terms of in-game accomplishments (and in complementary terms also the external pre-post-test for half of the participant group indicated that the children made progress in the domain). The different levels of difficulties that children with time would find themselves on did not seem to be perceived by the children as 'different levels'. They did engage in a variety of 'comparison talk' regarding their gardens, novel kinds of flowers, the sub-games played, etc. but no comparisons that related to how far they actually had progressed in terms of early mathematics. There were no indications that a child perceived herself as being

less capable than others in playing the game or as being behind—or ahead—of others.

In more general terms, the results from the study indicate that it is possible to design early math games that are both adaptive and inclusive.

Discussion on adaptivity and inclusion

The approach that MG uses to achieve *inclusion* can be compared to another approach that is often used as well as recommended: let all tasks be freely available for all children to choose from. With this approach, all children make their own choices which play down the exposure that can occur when a teacher makes the choices for different children. A drawback is, however, that young learners may not be capable of choosing appropriately challenging materials and tasks for themselves. Some prefer to remain in their comfort zone, and others take on tasks far too difficult for them.

Well-designed adaptive educational software, on the other hand, handles this since the choices made by the system are made on a pedagogical basis. An additional advantage with adaptive software compared to when a teacher chooses different tasks for different children to adapt to their differences is that the computer has no preconceptions (or prejudices). Teachers often come with different expectations on different children, based (more or less) on past realities—but the expectations also influence what *comes to be reality*. Children implicitly presented with low expectations (e.g. “if you can do these two first pages, that’s fine [sufficient for you]”) risk to align with these (Rattan et al. 2012). The computer, in contrast to the teacher, presents no expectations of this kind.

Adaptivity as goal for early math games is hardly controversial. However, the goal of *inclusion*—designing such games that prevent comparisons between children themselves with respect to (early) math competence—can be problematized, since competition can be used in games to motivate. Our take is that even though competition can be a source of motivation for learning, the costs in the domain of early math are larger than the gains. A young child who has been disadvantaged with respect to exposure to early math is, if competition is introduced at a time when she had had no chance to catch up, at risk to establish the false belief that she is less capable than others. Early math games should counteract the establishment of ill-grounded, low self-efficacy beliefs in preschoolers and young school children.

Limitations and future research

The lack of a control group where participants would have conducted the same external pre- and post-tests as 21 of the children in the intervention group is an obvious study limitation. Results from such a control group would clarify to what extent an improvement between the pre- and post-test has to do with general development. The relatively short intervention time of 6 weeks, however, decreases the likelihood

of general development being a primary explanation for the progress between pre- and post-test seen in the intervention group.

More importantly, the use of a control group could clarify to what extent the very experience of taking a test (the pre-test) leads to a better result the second time (the post-test) for participating children in this context. Therefore, follow-up studies should strive to include a control group.

In addition, the number of participants should preferably be larger in future studies. Another limitation of the study is that only children from middle-class communities participated. A follow-up study should include children from families of low socioeconomic *status*.

Above we discussed the fact that many early educational games exploit explicit references to progress in terms of achievement: stars, tokens, etc. corresponding to the levels of difficulties that have been accomplished. A future study could investigate in detail the extent to which, and how, children use such features to compare between themselves. So far, our observations when spending time at preschools indicate that it is common that children compare their achievements using these and similar features in early math software. And as reported in the present study, we observed that children extensively tried to do this with MG as well, although they could not really manage to find out how to compare.

To finish, a cultural note is in place. The concept of inclusion reported in this paper is very well received by teachers and parents in Sweden. It is an open question what the case is in other countries, given that views of competition in educational settings may differ. We look forward to future cross-cultural studies in this area.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43545-021-00205-7>.

Acknowledgements The presented research has been funded by The Wallenberg Foundations and by The Linneaus Research Environment 'Cognition, Communication and Learning' (via the Swedish Research Council).

Funding Open access funding provided by Lund University.

Data availability The data are available from the corresponding author upon reasonable request.

Declarations

Conflicts of interest There are no conflicts of interest or competing interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aleven V (2015) A is for Adaptivity, but what is Adaptivity? Re-Defining the field of AIED. In: Porayska-Pomsta K, McCalla G and du Boulay B (eds) Workshop Proceeding of AIED 2015 Vol 4: Workshop on Les Contes du Mariage: Should AI stay married to Ed? Madrid, pp 11–20
- Aubrey C, Dahl S, Godfrey R (2006) Early mathematics development and later achievement: further evidence. *Math Educ Res J* 18:27–46
- Bandura A (1997) Self-efficacy: the exercise of control. Freeman, New York
- Batchelor S, Torbeyns J, Verschaffel L (2019) Affect and mathematics in young children: an introduction. *Educ Stud Math* 100(3):201–209
- Blair KP (2013) Learning in critter corral: evaluating three kinds of feedback in a preschool math game. In: Sawhney N, Reardon E, Hourcade JP (eds) Conference Proceeding of IDC 2013. ACM, New York, pp 372–375
- Bransford J, Brown A, Cocking R (1999) How people learn: brain, mind, experience, and school. The National Academies Press, Washington, DC
- Bullough RV Jr, Hall-Kenyon KM, MacKay KL, Marshall EE (2014) Head start and the intensification of teaching in early childhood education. *Teach Teach Educ* 37:55–63
- Butterworth B, Laurillard D (2010) Low numeracy and dyscalculia: Identification and intervention. *ZDM Math Educ* 42:527–539
- Biswas G, Katzlberger T, Brandford J, Schwartz D, TAG-V (2001) Extending intelligent learning environments with teachable agents to enhance learning. In: Moore J, Redfield C, Johnson W (eds) Conference Proceeding of AIED 2001. IOS Press, Amsterdam, pp 389–397
- Chin DB, Dohmen IM, Schwartz DL (2013) Young children can learn scientific reasoning with teachable agents. *IEEE Trans Learn Technol* 6(3):248–257
- Cleary T, Zimmerman B (2012) A cyclical self-regulatory account of student engagement: theoretical foundations and applications. In: Christenson S, Reschly A, Wylie C (eds) Handbook of research on student engagement. Springer, Boston, pp 237–257
- Clements DH, Sarama J (2008) Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *Am Educ Res J* 45:443–494
- Deci E, Ryan R (1985) Self-determination and intrinsic motivation in human behavior. Kluwer Academic/Plenum Publishers, New York
- Denton K, West J (2002) Children's reading and mathematics achievement in kindergarten and first grade (Technical Report No NCES 2002125). National Center for Education Statistics, US Department of Education, Washington, DC
- Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P et al (2007) School readiness and later achievement. *Dev Psychol* 43:1428–1446
- Florian L, Spratt J (2013) Enacting inclusion: a framework for interrogating inclusive practice. *Eur J Spec Needs Educ* 28:119–135
- Ginsburg H, Carpenter K, Labrecque R (2011) Introduction to MathemAntics: Software for children from age 3 to grade 3. Paper presented at ISDDE 2011. Boston
- Ginsburg HP, Jamalian A, Creighan S (2013) Cognitive guidelines for the design and evaluation of early mathematics software: the example of MathemAntics. In: English L, Mulligan J (eds) Advances in mathematics education: reconceptualising early mathematics learning. Springer, Dordrecht, pp 83–120
- Griffin S, Case R, Siegler R (1994) Rightstart: providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In: McGilly K (ed) Classroom lessons: Integrating cognitive theory and classroom practice. The MIT Press, Cambridge, pp 24–49
- Griffin S (2007) Early intervention for children at risk of developing mathematical learning difficulties. In: Berch DB, Mazzocco MMM (eds) Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities. Paul H Brookes Publishing, Baltimore, pp 373–395
- Gulz A, Haake M (2019) Can preschoolers learn by teaching a digital character? (in Swedish). In: Riddersporre B, Kjällander S (eds) Digitalization in preschool on a scientific basis. Stockholm, Sweden, Natur och Kultur, pp 158–175
- Haake M, Axelsson A, Clausen-Bruun M, Gulz A (2015) Scaffolding mentalizing via a play-&-learn game for preschoolers. *Comput Educ* 90:13–23

- Husain L, Gulz A, Haake M (2015) Supporting early math – rationales and requirements for high quality software. *J Comput Math Sci Teach* 34(4):409–429
- Jordan NC, Kaplan D, Ramineni C, Locuniak MN (2009) Early math matters: kindergarten number competence and later mathematics outcomes. *Dev Psychol* 45(3):850–867
- Jordan NC, Glutting J, Dyson N (2012a) Number sense screener™(NSS™) User's guide, k–1 research editions. Brookes Publishing, Baltimore
- Jordan NC, Glutting J, Dyson N, Hassinger-Das B, Irwin C (2012b) Building kindergartners' number sense: a randomized controlled study. *J Educ Psychol* 104(3):647–660
- Klingner J, Vaughn S (1999) Promoting reading comprehension, content learning, and English acquisition through collaborative strategic reading (CSR). *Read Teach* 52(7):738–747
- Lee J, Park O (2008) Adaptive instructional systems. In: Spector JM, Merrill MD, Merriënboer JV, Driscoll MP (eds) *Handbook of research on educational communications and technology*, 3rd edn. Lawrence Erlbaum Associates, New York, pp 469–484
- Lepola J, Hannula-Sormunen M (2019) Spontaneous focusing on numerosity and motivational orientations as predictors of arithmetical skills from kindergarten to grade 2. *Educ Stud Math* 100(3):251–269
- McLeod D (1992) Research on affect in mathematics education: a reconceptualization. In: Grouws D (ed) *Handbook of research on mathematics learning and teaching*. Macmillan, New York, pp 575–596
- Missall K, Mercer S, Martínez R, Casebeer D (2012) Concurrent and longitudinal patterns and trends in performance on early numeracy curriculum-based measures in kindergarten and first grade. *Assess Eff Interv* 37(2):95–106
- Mononen R, Aunio P, Koponen T, Aro M (2014) A review of early numeracy interventions for children at risk in mathematics. *Int J Early Child Spec Educ* 6(1):25–54
- Pearce N, Conlan O, Wade V (2008) Adaptive educational games: Providing non-invasive personalised learning experiences. In: Eisenberg M, Kinshuk, Chang M, McGreal R (eds) *Proceeding of the DIGITEL 2008*. IEEE Computer Society, Los Alamitos, pp 28–35
- Praet M, Desoete A (2014) Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: a randomised controlled study. *Teach Teach Educ* 39:56–65
- Rattan A, Good C, Dweck C (2012) "It's ok – not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *J Exp Soc Psychol* 48(3):731–737
- Roos H (2019) Inclusion in mathematics education: an ideology, a way of teaching, or both? *Educ Stud Math* 100:25–41
- Schacter J, Shih J, Allen C, DeVaul L, Adkins A, Ito T, Jo B (2016) Math Shelf: a randomized trial of a prekindergarten tablet number sense curriculum. *Early Educ Dev* 27(1):74–88
- Sampayo-Vargas S, Cope C, He Z, Byrne G (2013) The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Comput Educ* 69:452–462
- Shute V, Zapata-Rivera D (2012) Adaptive educational systems. In: Durlach P, Lesgold A (eds) *Adaptive technologies for training and education*. Cambridge University Press, New York, pp 7–27
- Vanbecelaere S, Cornillie F, Sasanguie D, Reynvoet B, Depaepe F (2021) The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, non-cognitive and efficiency outcomes. *Br J Educ Technol* 52(1):112–124
- Vygotski L (1978) *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge
- Wilson A, Revkin S, Cohen D, Cohen L, Dehaene S (2006) An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behav Brain Funct* 2(1):1–16