The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, non-cognitive and efficiency outcomes

Stefanie Vanbecelaere[2,3*], Frederik Cornillie[3], Delphine Sasanguie[1], Bert Reynvoet[1], & Fien Depaepe[2,3]

[1] Brain & Cognition, KU Leuven Kulak, Kortrijk, Belgium

[2] Centre for Instructional Psychology and Technology, KU Leuven Kulak, Kortrijk, Belgium

[3] ITEC, KU Leuven, Belgium, also at imec Kapeldreef 75, B-3001 Leuven, Belgium

Author note

*Corresponding author. Faculty of Psychology and Educational Sciences, KU Leuven Kulak, Etienne Sabbelaan 51, 8500 Kortrijk, Belgium. E-mail address: stefanie.vanbecelaere@kuleuven.be

A set of structured practitioner notes

What is already known:

- Digit comparison and number line estimation are important predictors for the development of math.
- Digital game-based learning is effective in fostering early numerical skills.
- Empirical evidence on the effectiveness of adaptive compared to non-adaptive games is limited.

What this paper adds:

- Results revealed no significant differences between the adaptive and non-adaptive conditions for numerical ability, math achievement and math anxiety scores.
- Children in the adaptive condition learned more efficiently compared to the non-adaptive condition.
- It depends on children's prior knowledge how much they benefit from a (non-)adaptive training in terms of learning efficiency.

Implications for practice and/or policy:

- Adaptive games provide exercises according to the ability of the learner and are effective in fostering children's early numerical abilities.

Biography

The authors are part of the Faculty of Psychology and Educational Sciences at the KU Leuven, campus Kulak. Stefanie Vanbecelaere is a PhD student. Her research interests include the effectiveness of adaptive learning of math and reading. Frederik Cornillie is research manager of ITEC (research group of imec) at the campus Kulak of the University of Leuven. He coordinates projects within the field of digital innovation in education. Delphine Sasanguie is a postdoctoral researcher of the Fund for Scientific Research (FWO) and she is affiliated to the KU Leuven, Research unit Brain and Cognition, and Numerical Cognition lab in particular. Bert Reynvoet is full professor at the Faculty of Psychology and Educational Sciences and is affiliated to the Research unit Brain and Cognition. He is the head of the numerical cognition laboratory where developmental, experimental and neuroscience techniques are combined to unravel the cognitive underpinnings of calculation and numerical processes. Fien Depaepe is full professor at the Faculty of Psychology and Educational Sciences and is affiliated to the Research unit Center for Instructional Psychology and Technology. She is also a principal investigator of the imec KU Leuven interdisciplinary research group ITEC. Her main research interest deal with the instructional design of technology-enhanced learning environments.

Abstract

Adaptive educational games provide new opportunities to train early numerical skills. However, empirical evidence for the effectiveness of adaptive educational games is scarce. This study investigated the effectiveness of an adaptive game compared to a non-adaptive game in terms of cognitive, non-cognitive and efficiency outcomes. In total, 84 children were randomly assigned to a condition in which children trained early numerical skills with an adaptive version of the Number Sense Game (NSG), or to a condition in which they trained with a non-adaptive version. Early numeracy was evaluated before the training, immediately after the training and three weeks after the training. Math anxiety (MA) was assessed before and three weeks after the training. The time children practiced with the NSG was used to assess efficiency. Results revealed that children in both conditions improved on early numerical ability, with sustained effects three weeks after the training. In both conditions, children's MA scores were lower after the training. Children in the adaptive condition learned more efficiently compared to the non-adaptive condition, and the interaction between prior knowledge and condition has shown that children with low prior knowledge benefited more from a non-adaptive training while children with high prior knowledge benefited more from an adaptive training in terms of learning efficiency. These results confirm that adaptive educational games can offer solace in terms of the need for differentiation.

Keywords: adaptive learning, digital game-based learning, math anxiety, efficiency, individual differences

1 Introduction

An important goal of education is math development, which is influenced by cognitive and non-cognitive variables. Regarding cognitive factors, early numerical abilities are important and it has been suggested that training at a young age reduces math difficulties in children's future school careers (Schneider et al., 2017, 2018). Two important domain-specific tasks to train early numerical abilities are the magnitude comparison task and the number line estimation (NLE) task. Math achievement is also impacted by non-cognitive factors. For instance, prior research has revealed that math anxiety (MA) is negatively associated with math performance. These negative effects of MA can already appear in young children and can persist until adulthood (Ramirez, Shaw, & Maloney, 2018).

Innovative technologies such as digital game-based learning (DGBL) are specifically designed for educational purposes (Wouters, van Nimwegen, van Oostendorp, & van Der Spek, 2013) and have revealed new opportunities to foster early numerical abilities. Previous research has shown that specific features of digital games might enhance learning, such as automatically generated exercises, immediate feedback (O'Rourke, Main, & Hill, 2017), interaction between the system and the learner, concrete representations, and an attractive narrative (Ronimus, Kujala, Tolvanen, & Lyytinen, 2014). Recently, there has been increased attention for the integration of adaptivity which facilitates the alignment of exercise difficulty to the child's ability (Hooshyar, et al., 2019; Vanbecelaere et al., 2019).

Given these features, several scholars have stressed the possibilities of DGBL to enhance cognitive, non-cognitive, and efficiency learning outcomes (All, Núñez Castellar, & Van Looy, 2015; Wouters et al., 2013). Cognitive learning outcomes refer to academic outcomes such as math and reading and are typically evaluated with standardized tests. Examples of variables reflecting non-cognitive outcomes are motivation, attitudes or anxiety towards a certain topic which among others can be evaluated using questionnaires (Jansen et al., 2013; Van Roy & Zaman, 2017). Efficiency outcomes include the amount of time students need to gain mastery status (Kulik & Kulik, 1991). However, how these efficiency outcomes should be operationalized, is still debated (Hubalovsky, Hubalovska & Musilek, 2019). Empirical research investigating the effectiveness of DGBL has revealed mixed findings (Hooshyar et al., 2016; Jansen et al., 2013; Núñez Castellar et al., 2014; Núñez, Van Looy, Szmalec, & De Marez, 2014; O'Rourke et al. 2017; Ramirez et al., 2018; Van Roy & Zaman, 2017; Wouters et al., 2013).

A possible reason for the mixed findings is that research investigating the effectiveness of DGBL is characterized by several methodological shortcomings. First, most studies look at the overall effect of DGBL, although interventions might not be as effective for all learners (Miller & Robertson, 2011). For instance, Miller and Robertson (2011) observed larger learning gains of DGBL for students with lower cognitive abilities. Second, it is argued that comparing the learning gain of children who are enrolled in different versions of one educational game, is better suited to unravel which specific game features foster learning, instead of comparing the learning gains of educational games compared to a no game condition (Clark, Tanner-Smith, & Killingsworth, 2016).

2 Theoretical background

The implementation of adaptivity has been put forward as a promising feature to train particular skills more effectively and to reduce math anxiety (Klinkenberg, Straatemeier, & Van Der Maas, 2011; Sampayo-Vargas, Cope, He, & Byrne, 2013; Vanbecelaere et al., 2019). A learning environment is adaptive "if it is capable of monitoring the activities of its users, interpreting these on the basis of domain-specific models, inferring user requirements and preferences out of the interpreted activities, appropriately representing these in associated models; and finally, acting upon the available knowledge on its users and the subject matter at hand, to dynamically facilitate the learning process" (Paramythis & Loidl-Reisinger, 2004, p. 182). This study focuses on the adjustment of the content relative to the player's ability, which is one possibility to establish adaptivity. Below, an overview is provided of the effectiveness of adaptive games in terms of cognitive, non-cognitive and efficiency outcomes. As only few studies explicitly contrasted an adaptive with a non-adaptive game condition, we will also report studies comparing adaptive learning environments with pen-and-paper condition or passive control condition as well as studies outside the domain of early mathematics.

Cognitive factors

It is assumed that training according to children's ability will improve their learning. Hooshyar, Yousefi and Lim (2018) conducted an intervention study with 150 children being assigned to an adaptive or non-adaptive game condition to train phonological awareness and letter knowledge. The adaptive condition led to better performance on phonological awareness and letter knowledge. Similar effects were observed in an intervention study by Sampayo-Vargas et al. (2013) in which Spanish cognates were trained. The children were assigned either to an adaptive, non-adaptive or control condition. The results revealed that children in the adaptive condition outperformed children from the other conditions after training. However, the previous studies only assessed the trained skills immediately after training. It would be interesting to also include a delayed test to control for short-term novelty effects (Wouters et al., 2013). This is especially important in research comparing the effects of an adaptive with a non-adaptive condition as van Oostendorp, van der Spek and Linssen (2014) argued "in the adaptive game version, some children will receive less practice and consequently less opportunity to internalize the information. Therefore there is a real possibility – or even danger – that the participants in the adaptive condition remember less of the instruction after several weeks." (p. 7). In addition, it is also interesting to investigate whether the trained skills are transferred solely to tasks similar to the ones being trained in the game (i.e., the so-called 'near transfer'), or whether they are also generalized to other tasks (i.e., 'far transfer') (Cohen Kadosh, Dowker, Heine, Kaufmann, & Kucian, 2013).

Non-cognitive factors

Several researchers claim that early interventions should also try to reduce distress at failure, and preserve confidence in ability to solve math problems (Dowker, Cheriton, Horton, & Mark, 2019). Adaptive games might reduce frustrations, hence possibly reducing anxiety as children experience more success and less failure due to the integrated adaptivity (Hubalovsky et al., 2019; Jansen et al., 2013; Zohaib, 2018). Only a limited number of intervention studies investigated this assumption. Jansen and colleagues (2013) assigned children to an adaptive program for practicing math or to a control condition in which children practiced math as usual. From pretest to posttest, MA was reduced equally in all conditions. In a study by Núñez Castellar and colleagues (2014), young children were assigned to an adaptive game condition, a pen-and-paper condition or a passive control condition. Again, no differences were observed between the conditions.

Efficiency measures

Efficiency, considered as "reducing the time needed to teach a certain subject matter, resulting in similar learning outcomes" (All et al., 2015, p. 34) is a desired outcome of DGBL. Since adaptive games take into account individual differences between children they might contribute to efficiency learning outcomes (Hubalovsky et al., 2019). Quick learners are provided with less exercises and consequently, they need less time to achieve certain learning goals, whereas slower learners are presented with more exercises and will need more time to learn the same content (van Oostendorp et al., 2014). In a recent study, Hubalovsky and colleagues (2019) compared an adaptive with a non-adaptive e-learning environment on the efficiency of 52 children's learning process. The adaptive condition showed larger efficiency scores compared to the non-adaptive condition. Van Oostendorp and colleagues (2014) assigned students to an adaptive or a non-adaptive condition training triage (medical first response). Learning efficiency was measured by dividing learner performance by the number of cases triaged. Results revealed that participants in the adaptive condition learned significantly more per victim case, and were therefore more efficient than participants of the non-adaptive condition. Ali and Sah (2017) used a within subjects design to evaluate efficiency when learning geometry, physics and chemistry in an adaptive and non-adaptive game. Students showed higher efficiency scores in the adaptive compared to the non-adaptive game. However, so far, only a small number of studies took into account efficiency outcomes (van Oostendorp et al., 2014).

3 This study

The objective of this study is to examine the effectiveness of an adaptive and non-adaptive game on young children's learning. A core subject in the educational development of young children is math. As research has revealed that it is important to target the most powerful predictors in a domain (Duncan et al., 2007), our intervention focused on early numerical abilities through number comparison and NLE tasks (Schneider et al., 2017, 2018). Our study design addresses three shortcomings in the domain of DGBL. First, instead of comparing a game condition to a no game condition, we contrast an adaptive and a non-adaptive version of a game on children’s math performance to enhance our understanding of how adaptivity fosters learning. Second, the effectiveness of adaptivity is investigated in a broad sense: in terms of cognitive, non-cognitive, and efficiency outcomes (All et al., 2015). For the cognitive outcomes, near transfer (digit comparison, NLE) and far transfer (pictorial calculating, simple arithmetic) are evaluated immediately after the intervention and delayed after three weeks. Concerning non-cognitive learning outcomes, children’s anxiety towards math is assessed. Efficiency is examined using performance scores and how much time children spent in the game. Third, not only overall effectiveness is examined but also individual differences are taken into account. The following research questions are addressed: (1) What is the effectiveness of adaptive and non-adaptive digital games on cognitive outcomes? More specifically on (a) immediate and (b) delayed cognitive outcomes; (2) What is the effectiveness of adaptive and non-adaptive digital games on non-cognitive outcomes?, (3) What is the effectiveness of adaptive and non-adaptive digital games on efficiency outcomes?, (4) How does prior knowledge moderate the intervention effects on efficiency outcomes?.

4 Methodology

Number sense game

We used the number sense game (NSG) (Linsen et al., 2015) consisting of a comparison and a NLE game (see Figure S1). In the comparison game, the levels are designed to vary in difficulty based on the numerosities (1-4, 1-9, 5-18), the display duration and the type of stimuli (non-symbolic, symbolic and mixed notation). The difficulty of the levels in the NLE game depends on the number of anchor points, the display duration, and the type of stimuli (non-symbolic, symbolic, and mixed notation).

Description of the conditions

A non-adaptive and adaptive version of the NSG were developed differing only in the difficulty adjustment mechanism used. In the non-adaptive NSG, the difficulty level increases based on a pre-defined series of exercises. Regardless of the child's performance in the game, the child proceeded to the next level after having completed the tasks of the current level. Consequently, how much time children needed to finish the game depended on their reaction time for each item as the number of exercises and the order of the items was exactly the same for all children. A detailed overview of the levels in the game can be found in Linsen and colleagues (2015, pp. 14-22).

In the adaptive NSG, psychometric modelling techniques based on the Elo-rating algorithm were used to incrementally build user and content models (Klinkenberg et al., 2011). The simplest IRT model, the Rasch model, states that the probability of a correct answer depends on the person's ability and the item's difficulty. Both kinds of parameters can be estimated from participants' responses, and placed on a common scale. Therefore data of a previous study were used and IRT-analysis was conducted (Vanbecelaere et al., 2020). During gameplay, the Elo-rating system was used to update the person's ability estimates only, while the difficulty parameters were considered as constants over time. When a learner finishes a level, based on his/her current ability, an appropriate level is presented. In practice, children with weaker abilities progress more slowly because levels of the same difficulty sometimes need to be repeated. Strong players run through the games more quickly as they are allowed to skip levels.

Procedure

In total, 88 1st graders from five classes of three different elementary schools participated. All children were randomly (i.e., using individual randomized trial) assigned to two experimental conditions. The non-adaptive NSG condition consisted of 39 children, whereas the adaptive NSG contained 45 pupils. Participants assigned to the adaptive or non-adaptive condition were not informed to which condition they belonged. Pupils were excluded (*n*=6) in case their parents did not sign the informed consent, they were absent during one of the NSG-interventions, or they did not finish the NSG after six consecutive training sessions. All children of the participating classes took part in the study, but in case parents did not give consent, the logging data and the results were not used for analysis and all data related to these participants were removed. Analyses were conducted on the data of the remaining 78 children.

Pretests were administered one week before the intervention. The intervention consisted of 30-minutes training per week over a period of three weeks resulting in six training sessions. In each session, a researcher went to the classroom to distribute the tablets and headphones and supported the pupils during their individual play together with their regular teacher(s). Children started with playing the digit comparison game and when finished, continued with the NLE game. After finishing both games, immediate posttests were administered. After three weeks the delayed posttests were taken (see Figure S2).

Materials

4.1.1 Cognitive factors

To examine near and far transfer effects, several validated instruments were used. Near transfer tasks are similar to the tasks presented in the game. In this study, the digit comparison task (Brankaer, Ghesquière & De Smedt, 2017) and NLE task (Siegler & Booth, 2004) were used (Appendix S3, S4). A detailed description of these tasks can be found in Vanbecelaere and colleagues (2020).

Far transfer tasks measure the transfer to general math outcomes. General math competence was assessed with the 'pictorial calculation' and 'simple arithmetic' subtest of the TEDI-MATH (Grégoire, Noël, & Van Nieuwenhoven, 2004) (Appendix S5). Children conducted both subtests in a one-to-one setting with the experimenter. The pictorial calculation subtest consisted of 6 items (e.g. Here you see two red balloons and three blue balloons. How many balloons are there together?"). The experimenter reads the questions one by one, and the child responded orally. For each correct answer, the child was given one point. In this study, the internal consistency of this subtest is rather weak (Cronbach's alpha = .63). The simple arithmetic subtest consisted of 18 exercises with increased difficulty. Children were given a booklet with on each page one exercise, and were asked to solve the exercises one by one by saying the correct answer aloud. In case of a correct answer, children were instructed to proceed. When children made five consecutive mistakes, the experimenter finished the test because it was assumed that from that point onwards, children would not be able to answer more difficult exercises. In this way we prevented children from being frustrated. The total score was the number of correctly answered problems.

4.1.2 Non-cognitive factors

MA was assessed with the Kinder-Angst-Test-III (KAT-III) (Orbach, Herzog, & Fritz, 2018; Tewes & Naumann, 2017)(Appendix S6). The items of the KAT-III were translated to Dutch. In a one-to-one setting with the experimenter and the child, the child was told that he/she would have to solve a math test, but before starting this test, they had to respond to some questions. The questionnaire consisted of 8 items and was administered before (e.g., "I'm nervous") and after (e.g., "I was nervous") the administration of the TEDI-MATH test. The children were asked to answer to the questions by means of pointing at 'yes', 'rather yes', 'rather no' or 'no'. The total MA score is calculated based on these 16 items. Reliability analysis of the questionnaire during pretest was rather weak and sufficient at the time of the posttest (Cronbach's alpha = resp. .62 and .72).

4.1.3 Efficiency

During the intervention the time spent on digit comparison and NLE tasks was registered to determine whether learning was more efficient in the adaptive compared to the non-adaptive condition. When children saw the final generics of the digit comparison game and thus finished, they were instructed to immediately inform the experimenter who subsequently registered how long the child needed to finish this game. After registering the time, the experimenter started the second game training NLE. Similarly, when a child finished, the time they spent on the NLE training was again registered by the experimenter.

Data-analysis

The assumptions for the analyses we have conducted were met[1]. The first research question examined the effects of the adaptive and non-adaptive NSG game on cognitive outcomes. In view of determining near transfer effects (i.e. digit comparison, NLE) immediately after the intervention, we ran two Repeated Measures Analyses of Variance (ANOVA) with test moment (2 levels: pretest versus posttest) as within-subjects variable and condition (adaptive versus non-adaptive condition) as between-subjects factor. For digit comparison, motor speed was included as covariate. Furthermore, delayed effects for near transfer (i.e. digit comparison, NLE) and far transfer (i.e. pictorial calculation, simple arithmetic) were examined through four Repeated Measures ANOVAs with test moment as within-subjects variable and condition as between-subjects factor. For digit comparison, motor speed was included as covariate.

The second question investigated the impact of adaptive and non-adaptive games on MA. A repeated-measures ANOVA was conducted with moment as within-subject factor and condition as between-subjects factor.

[1]We did normality checks and used Levene’s test on the residuals of each model used in this study to answer the research questions. The model with pictorial calculation scores showed significant Kolmogorov-Smirnov test, but this was not problematic due to the fact that we compared groups of equal group sizes in each condition. For the model including ‘time DC’ as independent variable, the variances were not equal across the conditions. However in terms of violation, the test is fairly robust in terms of the error rate when sample sizes are equal (Field, 2013).

The third question addressed whether the adaptive condition learned more efficient compared to the non-adaptive condition. Two ANOVAs were conducted with respectively total time spent for digit comparison and total time spent for the NLE as dependent variables and condition as factor.

Fourth, to investigate whether prior knowledge moderated the intervention effects on efficiency outcomes, two separate ANCOVAs were ran. For the first analysis we included the pretest scores of digit comparison as covariate and condition was included as independent variable. Total time spent in the digit comparison game was used as dependent variable. The second analysis was ran with pretest scores of NLE as covariate, condition as independent variable and total time spent in the NLE game as dependent variable.

5 Results

Table S7 shows the means and standard deviations for each condition on the different outcome variables, i.e. pretest, posttest and delayed test of the cognitive, non-cognitive and total time spent measures. One-way ANOVAs on all pretest measures showed no differences between the two experimental conditions: digit comparison ($F(1,77)=0.34$, $p=.56$), NLE ($F(1,77)=0.70$, $p=.41$), pictorial calculation ($F(1,76)=0.17$, $p=.68$), simple arithmetic ($F(1,76)=0.02$, $p=.90$), MA ($F(1,73)=0.12$, $p=.73$). The following sections will describe the results of the three research questions.

Effects on cognitive outcomes

5.1.1 Immediate posttest

We observed a main effect of test moment, $F(1,75)=15.21$, $p=.00$, $\eta_p^2=.17$, for digit comparison. Children in both conditions significantly improved on this task. However, there was no test moment x condition interaction, indicating that children of both conditions improved to the same extent on the digit comparison task. Similar results were observed for the NLE task. Both conditions improved on the posttest, $F(1,76)=33.02$, $p=.00$, $\eta_p^2=.30$, but the interaction with group was not significant.

5.1.2 Delayed posttest

Regarding near transfer, a main effect of test moment was observed in the digit comparison task, $F(1,69)=9.34$, $p=.00$, $\eta_p^2=.12$, which means that children from both conditions performed better on the delayed test compared to the pretest. The results revealed no significant moment x condition interaction effect, indicating that children from both conditions improved to the same extend. For the NLE task the results are similar: a main effect of test moment, $F(1,71)=15.82$, $p=.00$, $\eta_p^2=.18$, but no significant interaction with condition, showing that both conditions improved to the same extent.

For far transfer, a significant main effect of moment was observed in the pictorial calculation task, $F(1,66)=25.30$, $p=.00$, $\eta_p^2=.28$, indicating that children performed better on the delayed test. The interaction between moment x condition was not significant. Similar results were obtained for the simple arithmetic task: i.e., a main effect of moment, $F(1,67)=34.40$, $p=.00$, $\eta_p^2=.34$, but no interaction with condition.

Effects on non-cognitive outcomes

For MA, a main effect of moment was observed, $F(1,65)=4.32$, $p=.04$, $\eta_p^2=.06$, revealing that MA was reduced after the intervention. Again, we did not observe an interaction of moment x condition.

Effects on efficiency outcomes

Regarding efficiency, we observed a significant difference between both conditions for total time spent (Table S8), both for digit comparison $F(1,77)= 555.42$, $p=.00$, $\eta_p^2=.88$, and NLE, $F(1,77)=7.56$, $p=.01$, $\eta_p^2=.09$. The non-adaptive condition trained significantly longer than the adaptive condition. Logdata confirmed that in the non-adaptive condition there was no correlation between the number of completed tasks and the time children needed to finish the game whereas in the adaptive condition a logic correlation was established indicating that children who needed less time completed less tasks (and vice versa). Because the performance on cognitive outcomes in the posttest did not differ between conditions, it can be concluded that the adaptive condition led to more efficient learning compared to the non-adaptive condition.

Further analyses were conducted to examine whether efficiency outcomes in the two conditions were moderated by prior knowledge. For digit comparison, results revealed a significant main effect of condition, $F(1,74)=110.47$, $p=.00$, $\eta_p^2=.60$, and a significant main effect of prior knowledge, $F(1,74)=12.44$, $p=.00$, $\eta_p^2=.14$. This means that children in the adaptive condition needed less time than the children in the non-adaptive condition. In addition, children with higher prior knowledge needed less time compared to children with lower prior knowledge. The interaction between condition and prior knowledge was not significant (Figure 1).

For NLE, results revealed a significant main effect of condition, $F(1,74)=22.89$, $p=.00$, $\eta_p^2=.24$, indicating that again children in the adaptive condition needed less training time. Furthermore, a significant main effect of prior knowledge was observed, $F(1,74)=18.59$, $p=.00$, $\eta_p^2=.20$, in favor of children with higher prior knowledge. Additionally, the interaction between condition and prior knowledge was significant, with $F(1,74)=15.39$, $p=.00$, $\eta_p^2=.17$, indicating that children with low/high prior knowledge needed different training time dependent on the condition to which they were belong. Figure 1 shows that in the non-adaptive condition, children with low or high prior knowledge on average spent the same amount of time. By contrast, in the adaptive condition, children with high prior knowledge spent less time in the game compared to low prior knowledge children. Furthermore, children with low prior knowledge needed significantly more time in the adaptive condition compared to low prior knowledge peers in the non-adaptive condition. By contrast, children with high prior knowledge needed significantly less time in the adaptive condition compared to high prior knowledge peers in the non-adaptive condition.

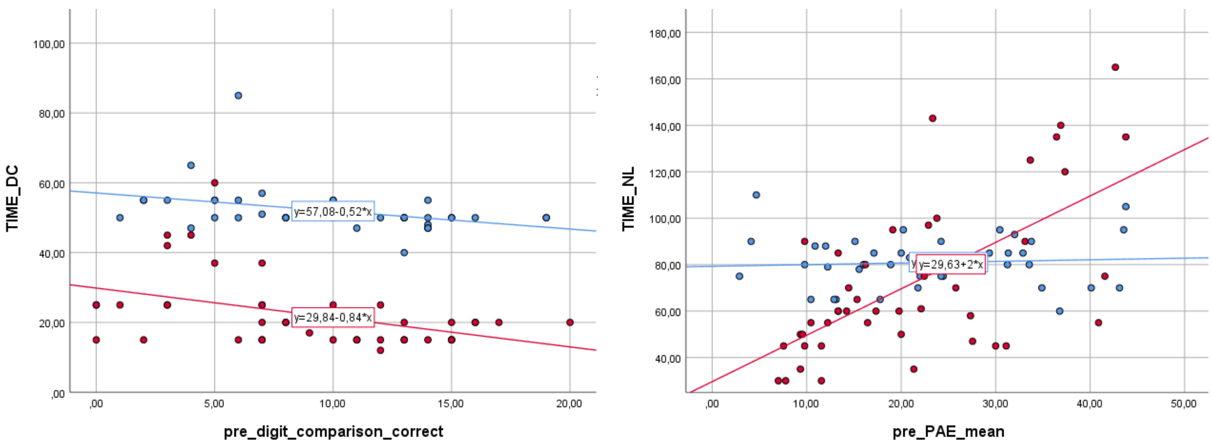


Figure 1. Scatter plot of the pretest scores (left: digit comparison, right: NLE) against the total time spent during the intervention (blue: non-adaptive condition, red: adaptive condition)

6 Discussion

In this study, children trained early numerical skills through a NSG consisting of digital comparison and NLE tasks. Children were randomly assigned to an adaptive and a non-adaptive condition. The effects of the intervention were investigated with respect to cognitive, non-cognitive and efficiency outcomes.

The first research question determined the effect of the (non-)adaptive game training on numerical abilities. Children improved significantly from pretest to posttest on near transfer tasks (i.e., digit comparison, NLE) in both conditions immediately after the intervention. However, both conditions were equally beneficial. This is not in line with previous research showing better learning outcomes for immediate near transfer for an adaptive compared to a non-adaptive condition (Núñez Castellar et al., 2014; Hooshyar et al., 2018; Sampayo-Vargas et al., 2013). As indicated by several meta-analyses and reviews (Nelson and McMaster, 2019; Reynvoet, Vanbecelaere, Sasanguie & Depaepe, accepted), these differences might be explained by the tests that were used (self-developed versus standardized), the targeted domain (e.g., math, language) or the operationalization of the non-adaptive condition (e.g., digital or with pen-and-paper).

Three weeks after the intervention, consolidation of the (non-)adaptive game training in terms of near (digit comparison and NLE) and far (pictorial calculation, simple arithmetic) transfer was examined. When interpreting the results, one should take into account the rather weak reliability of the pictorial calculating test. Children improved from pretest to delayed test on all measures. Next to a training effect, the learning gain might also result from regular learning activities during the time of the study. In addition, a test-retest effect could have played a role. As we did not make use of a control group, we are not sure that the observed cognitive learning gains are only due to the training effect. However, a previous study with the NSG showed that the experimental condition (non-adaptive version of the NSG) obtained similar results as the control condition for digit comparison, whereas for NLE the experimental condition outperformed the control condition after eight weeks of training (Vanbecelaere et al., 2020). Concerning the effect of condition on the delayed tests, no significant differences between the adaptive and non-adaptive condition were observed on near and far transfer tests. This is an interesting finding because some authors have concerns whether skipping exercises when developing lower cognitive domains (such as digit comparison, NLE) would influence the consolidation of the learning and/or the acquisition of more general math skills (Van Oostendorp et al., 2014). Our results are in line with the findings of Hubalovsky and colleagues (2019) and support the idea that the adaptivity algorithm might be able to make the decision instead of the learner to skip exercises when the ability of the learner is higher than the difficulty of the item during drill-and-practice training of early numerical skills.

In terms of non-cognitive learning outcomes, MA of children of both conditions significantly decreased after the intervention. Nevertheless, no significant difference between both conditions was observed. Consequently, our findings are not consistent with the prior hypothesis that adaptive game training would be able to reduce MA in young children (Jansen et al., 2013; Nunez Castellar et al., 2014). A possible explanation for this result is that differences between both versions of the NSG are too small to be noticed by the children. The reliability of the MA questionnaire was also rather weak. Furthermore, as we did not have any control group in the present study, we do not know whether a reduction of anxiety is due to game-based learning. Based on a previous study with the non-adaptive game and a non-game based control condition, we observed that math anxiety decreased to a similar extent in both conditions (Vanbecelaere et al., 2020).

Third, in terms of efficiency outcomes, the ‘total time spent’ outcome showed that children spent on average less time in the adaptive compared to the non-adaptive condition, both on digit comparison and NLE. As the results on the cognitive measures showed that children equally improved in both conditions, we can conclude that children learnt more efficiently in the adaptive condition. These results are in line with previous studies observing that learning in adaptive conditions is more efficient compared to non-adaptive conditions (Ali & Sah, 2017; Hubalovsky et al., 2019; Van Oostendorp et al., 2014).

Finally, we examined whether there were individual differences in efficiency outcomes based on prior knowledge, condition and their interaction. For digit comparison, the results revealed that children with low/high prior knowledge did not differ from each other as a result of the training in the adaptive or the non-adaptive condition. For NLE, children with low prior knowledge in the adaptive condition spent significant more time compared to children who played the non-adaptive condition. Children with high prior knowledge, played a lot faster in the adaptive condition, compared to those of the non-adaptive condition. This might be explained by the fact that digit comparison skills were already developed to a certain extent at the beginning of the training. All children improved, but children with low/high prior knowledge did not spend more or less time in one or the other training. For NLE, this content was quite new for all children. Consequently, children with low prior knowledge have spent more time in the adaptive compared to the non-adaptive condition, whereas children with high prior knowledge spent less time in the adaptive compared to the non-adaptive condition. Interestingly, it seems that low prior knowledge children learn more efficient in the non-adaptive condition, suggesting that low performing children benefit more from a pre-determined trajectory in which the exercises are structured with a pre-defined increased difficulty. Conversely, children with high prior knowledge benefit more from the adaptive condition in which they can advance faster, and thus learn more efficient.

Despite the merits of our study in terms of a comprehensive understanding of the effects of adaptive versus non-adaptive games, several questions remain unanswered. A first question relates to the operationalization of the non-cognitive outcomes. Although in this study math anxiety was included, it is assumed that these results also would apply to other non-cognitive measures such as motivation or attitudes since these concepts are distinct but related to each other (Baten, Vansteenkiste, De Muynck, De Poortere, & Desoete, 2019). Nevertheless, few empirical studies until today have shown mixed results regarding the effect of adaptive games on a range of different non-cognitive outcomes (Jansen et al., 2013; Núñez Castellar et al., 2014). Therefore, future work might include other non-cognitive measures to prove this assumption. Second, it was observed that previous research has adopted different methods to operationalize this learning efficiency so far (Ali & Sah, 2017; Hubalovski et al, 2019; Van Oostendorp et al., 2014). In this study, we separately analysed performance on posttest (research question 1) and total time spent (research question 3). An alternative way to measure efficiency is to divide posttest performance scores by the number of completed exercises (Van Oostendorp et al. 2014). Yet another approach is to use log data during learning. For example, learning efficiency can be calculated by dividing success of completed exercises by the participant's completion time (i.e. time needed for exercise completion; Hubalovsky et al., 2019). Finally, the study of Ali and Sah (2017) used performance scores during learning and the time participants needed to finish the quiz was logged and independently compared to examine whether participants were more efficient in one or the other condition. As also pointed by Hubalovsky and colleagues (2019), an interesting direction for future research would be to investigate how learning efficiency should be operationalized and to document the advantages and disadvantages of these methods.

Another question is whether children would obtain better learning results in the adaptive condition compared to the non-adaptive condition when they would have more time to spend in the learning game. Meta-analyses investigated whether the duration of the intervention moderates the effect of an intervention and showed that generally short interventions are more effective than long interventions (Nelson & McMaster, 2019; Reynvoet et al., accepted). However, long interventions typically target a broad range of skills whereas short trainings are more focused which explains this counterintuitive finding. In our study, we focused on two specific early numerical skills, and the end goals were reached in both conditions. By using a longitudinal intervention in which learners can continue with more difficult exercises, it is hypothesized that children in the adaptive condition not only would reach the same learning goals in less time compared to the children of the non-adaptive condition, but also that they would show better results on cognitive tests. We further believe that longer gameplay would improve the learning algorithm behind the adaptive software, and in this way increase the impact of adaptive educational games.

7 Conclusion

In sum, the adaptive and non-adaptive game condition did not differ in cognitive and non-cognitive learning outcomes, but significant differences were observed in terms of efficiency. Children who played the adaptive version spent significantly less time compared to children who played the non-adaptive version of the NSG. These results show that children learn with different paces, and that adaptive educational games can offer solace in terms of the need for differentiation.

8 Acknowledgements

This study was made possible by the imec-icon project LEAPS (Learning Analytics for AdaPtive Support) in which the research groups ITEC and mintlab (2 imec research groups at KU Leuven) and IDLab (an imec research group at UGent) collaborated with the industrial partners Pelckmans and Sensotec. The project was carried out within the imec's Smart Education research programme, co-financed by imec and received project support from VLAIO Flanders Innovation & Entrepreneurship.

9 Statements on potential conflicts of interest, open data, ethical guidelines and approval for reports of empirical research

The datasets generated during and/or analyzed during the current study are not publicly available due to human subject grounds but are available from the corresponding author on reasonable request.

Prior to the study, approval of the social and societal ethics committee was confirmed (G-2018 03 1194).

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

10 References

Ali, A., & Sah, M. (2017). Adaptive game-based e-leaming using semantic web technologies. In *2017 International Conference on Open Source Systems & Technologies (ICOSST)* (pp. 15-23). IEEE.

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2015). Towards a conceptual framework for assessing the effectiveness of digital game-based learning. *Computers and Education*, *88*, 29–37.

Baten, E., Vansteenkiste, M., De Muynck, G. J., De Poortere, E., & Desoete, A. (2019). How can the blow of math difficulty on elementary school children's motivational, cognitive, and affective experiences be dampened? The critical role of autonomy-supportive instructions. *Journal of Educational Psychology*.

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods*, *49*(4), 1361–1373.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: a systematic review and meta-analysis. *Review of Educational Research*, *86*(1), 79–122.

Cohen Kadosh, R., Dowker, A., Heine, A., Kaufmann, L., & Kucian, K. (2013). Interventions for improving numerical abilities: Present and future. *Trends in neuroscience and education, 2*(2), 85-93.

Dowker, A., Cheriton, O., Horton, R., & Mark, W. (2019). Relationships between attitudes and performance in young children's mathematics. *Educational Studies in Mathematics*, *100*(3), 211-230.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental psychology*, *43*(6), 1428.

Grégoire, J., Noël, M., and Van Nieuwenhoven, J. (2004). TEDI-MATH. Antwerpen: Harcourt.

Hooshyar, D., Ahmad, R. B., Yousefi, M., Fathi, M., Horng, S. J., & Lim, H. (2016). Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Computers & Education, 94*, 18-36.

Hooshyar, D., Lim, H., Pedaste, M., Yang, K., Fathi, M., & Yang, Y. (2019, December). AutoThinking: an adaptive computational thinking game. In *International Conference on Innovative Technologies and Learning* (pp. 381-391). Springer, Cham.

Hooshyar, D., Yousefi, M., & Lim, H. (2018). A procedural content generation-based framework for educational games: toward a tailored data-driven game for developing early English reading skills. *Journal of Educational Computing Research, 56*(2), 293–310.

Hubalovsky, S., Hubalovska, M., & Musilek, M. (2019). Assessment of the influence of adaptive E-learning on learning effectiveness of primary school pupils. *Computers in Human Behavior, 92*, 691-705.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24*, 190-197.

Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education, 57*(2), 1813-1824.

Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in human behavior, 7*(1-2), 75-94.

Linsen, S., Maertens, B., Husson, J., Van den Audenaeren, L., Wauters, J., Reynvoet, B., ... & Elen, J. (2015). Design of the game-based learning environment "Dudeman & Sidegirl: Operation Clean World," a numerical magnitude processing training. In *Describing and Studying Domain-Specific Serious Games* (pp. 9-26). Springer, Cham.

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology, 42*(5), 850-864.

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

Núñez Castellar, E., Van Looy, J., Szmalec, A., & De Marez, L. (2014). Improving arithmetic skills through gameplay: Assessment of the effectiveness of an educational game in terms of cognitive and affective learning outcomes. *Information sciences*, *264*, 19-31.

Orbach, L., Herzog, M., & Fritz, A. (2019). Relation of state- and trait-math anxiety to intelligence, math achievement and learning motivation. *Journal of Numerical Cognition. 5*(3), 371-399.

O'Rourke, J., Main, S., & Hill, S. (2017). Commercially available digital game technology in the classroom: Improving automaticity in mental-maths in primary-aged students. *Australian Journal of Teacher Education*, *42*(10), 50–70.

Paramythis, A., & Loidl-Reisinger, S. (2003). Adaptive learning environments and e-learning standards. In *Second european conference on e-learning* (Vol. 1, No. 2003, pp. 369-379).

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist, 53*(3), 145-164.

Reynvoet, B., Vanbecelaere, S., Sasanguie, D., & Depaepe, F. (accepted). Intervention studies in math: a discussion. In *Learning and education in numerical cognition*.

Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education, 71*, 237-246.

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers and Education*, *69*, 452–462.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, *20*(3).

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: a meta-analysis. *Child Development*, *89*(5), 1467–1484.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444.

Tewes, A., & Naumann, A. (2017). *Kinder-Angst-Test-III [Children-Anxiety-Test-III]*. Göttingen: Hogrefe.

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effects of two digital educational games on cognitive and non-cognitive math and reading outcomes. *Computers & Education*, *143*, 103680.

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2019). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*.

van Oostendorp, H., van der Spek, E. D., & Linssen, J. (2014). Adapting the complexity level of a serious game to the proficiency of players. *EAI Endorsed Transactions on Game-Based Learning*, *1*(2), e5.

van Roy, R., & Zaman, B. (2017). Why gamification fails in education and how to make it successful: introducing nine gamification heuristics based on self-determination theory. In *Serious Games and edutainment applications* (pp. 485-509). Springer, Cham.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265.

11 Supplementary Files

Figure S1

Screenshots of the comparison game (first picture) and the NLE game (second picture)

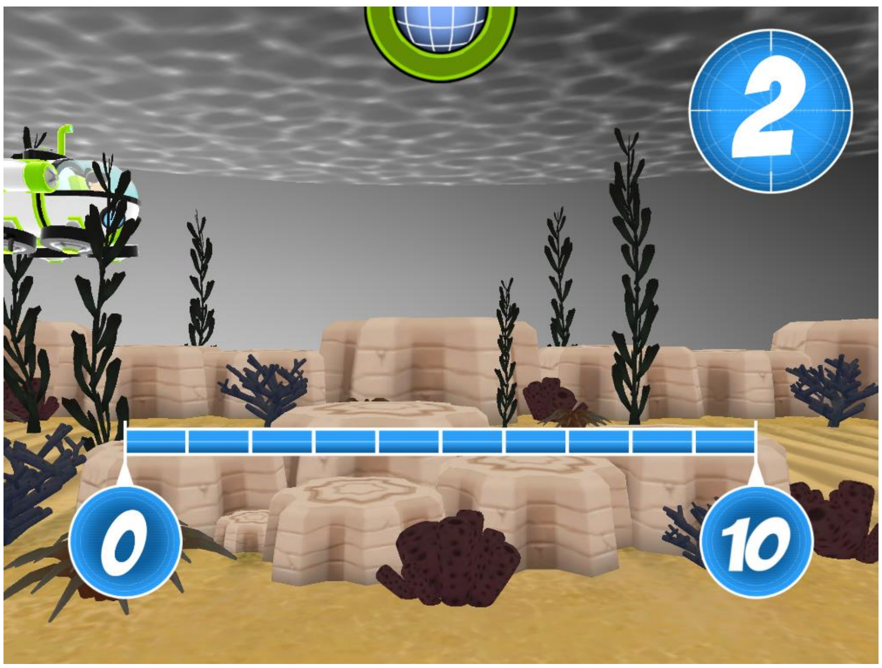
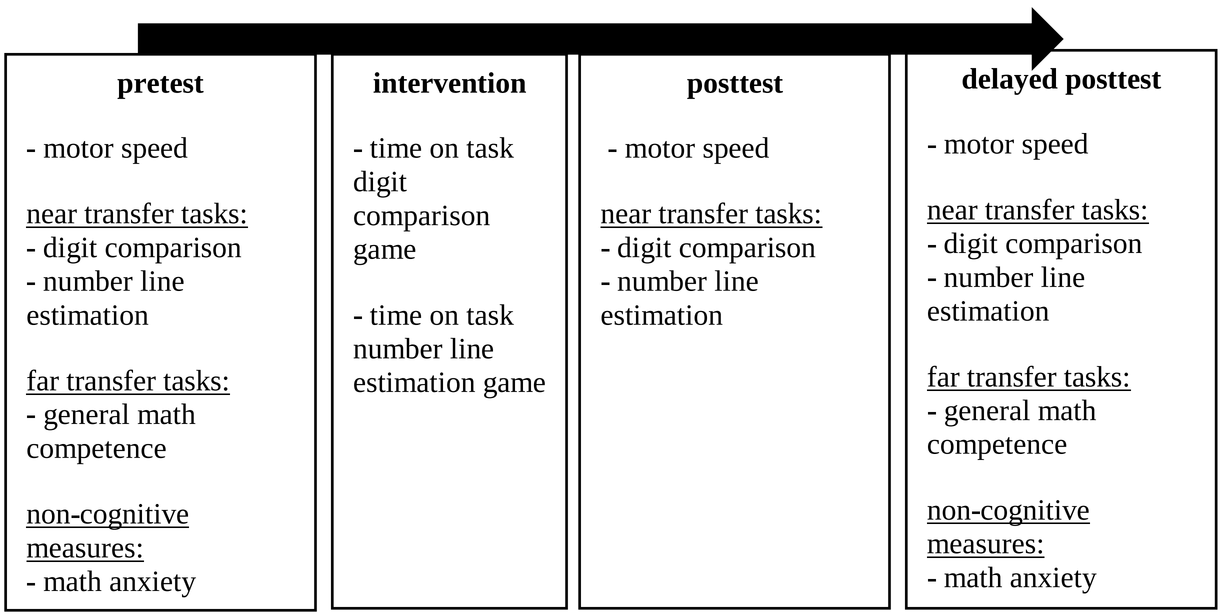


Figure S2

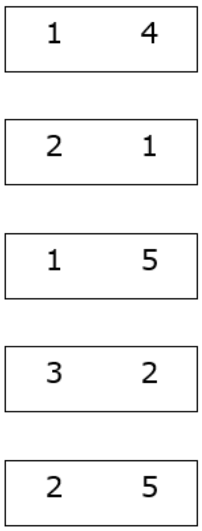
Design and materials



Appendix S3

Digit comparison

These are five examples items of the digit comparison task.



Appendix S4

Number line estimation

These are two examples of the number line estimation task.

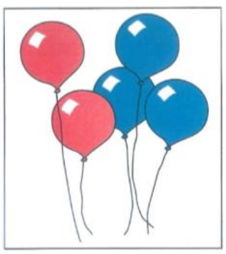
Appendix S5

General math competence

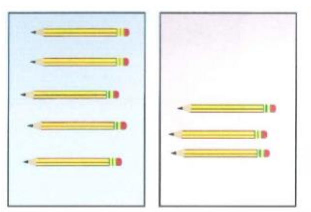
Pictorial calculation subtest

Examples of items:

- "Here you see two red balloons and three blue balloons. How many balloons are there together?"



- "There are five pencils on the blue sheet. On the pink sheet, you see three pencils. How many pencils you see in total?"



Simple arithmetic subtest

Instructions: Put the first paper (2+2) in front of the child and read: "two plus two, is equal to?". Show the child all items one by one (making use of a booklet) without reading aloud: 0+8, 6+3, 5+0, 3+5, 4+6, 7+7, 9+4, 6+8, 5+7, 20+8, 32+14, 20+30, 28+41, 24+18, 28+34, 45+16 and 35+17. The teacher says: Calculate every exercise in your head and tell me the answer. The experimenter stops the test when the child makes five consecutive mistakes.

Appendix S6

Math anxiety

Kinder-Angst-Test-III (KAT-III) (Orbach, Herzog, & Fritz, 2018; Tewes & Naumann, 2017)

Examples of items (before the test/after the test):

- I am excited / I was excited
- I am nervous / I was nervous
- I am worried / I was worried

Response possibilities:

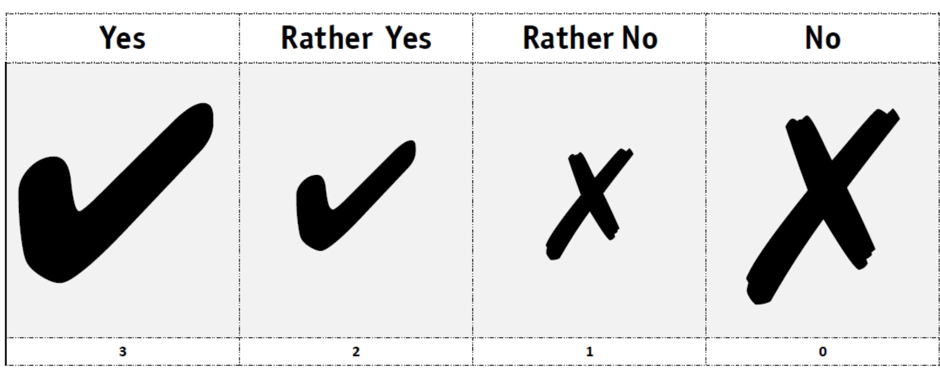


Table S7

Results on pretest, posttest and delayed test for cognitive and non-cognitive tests

test	condition	pretest			posttest			delayed posttest		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
motor speed	non-adaptive	37	10.73	4.34	37	14.41	3.87	34	14.65	3.91
	adaptive	41	11.59	4.68	41	12.41	4.09	44	14.50	4.42
digit comparison	non-adaptive	37	9.95	4.72	37	13.30	3.91	35	15.75	1.37
	adaptive	41	9.29	5.14	41	12.07	3.74	43	14.47	4.58
PAE Mean Scores	non-adaptive	37	22.69	11.66	37	14.84	7.72	35	17.77	8.83
	adaptive	41	20.59	10.46	41	14.23	7.35	43	15.80	7.11
pictorial calculation (max. 6)	non-adaptive	33	4.64	1.48	[a]	[a]	[a]	33	5.39	0.93
	adaptive	35	4.29	1.45	[a]	[a]	[a]	35	5.34	0.80
simple arithmetic	non-adaptive	37	3.41	3.56	[a]	[a]	[a]	33	4.64	3.14
(max. 17)	adaptive	40	3.30	3.89	[a]	[a]	[a]	36	4.61	3.47
math anxiety (max. 48)	non-adaptive	38	20,58	5,51	[a]	[a]	[a]	35	19,09	1,24
	adaptive	41	19,41	6,87	[a]	[a]	[a]	40	17,38	1,09

Table S8

Results for learning gain and total time spent

variable	condition	Results		
		N	**Mean**	**SD**
total time spent (DC)	non-adaptive	37	50.84[a]	3.30
	adaptive	41	20.68[a]	7.12
total time spent (NLE)	non-adaptive	37	80.78[a]	11.65
	adaptive	41	67.15[a]	28.06

[a] expressed in minutes