Title

The effect of adaptivity in digital learning technologies. Modeling learning efficiency using data from an educational game.

Authors

Dries Debeer[*1,2], Stefanie Vanbecelaere[*1,2], Wim Van den Noortgate[1,2], Bert Reynvoet[1], Fien Depaepe[1,2]

1 KU Leuven, Faculty of Psychology and Educational Sciences

2 KU Leuven, imec research group itec

*these authors contributed equally to this work.

Corresponding author: Dries Debeer, KU Leuven campus Kulak Kortrijk, E. Sabbelaan 51, 8500 Kortrijk, Belgium

E-mail address: dries.debeer@kuleuven.be

URL: https://itec.kuleuven-kulak.be/?page_id=5357

A set of structured practitioner notes

What is already known about this topic:

- Adaptive digital technologies are able to address students' individual needs
- Pretest posttest intervention research on the effectiveness of adaptive digital technologies revealed mixed results

What this paper adds:

- Learning effectiveness is different from learning efficiency
- Learning efficiency can be operationalized using log-data to model students' progress over time
- Adaptive digital technologies increased students' learning efficiency compared to non-adaptive digital technologies

Implications for practice and/or policy:

- Log-data can be obtained unintrusively and provides fine-grained information about students' learning
- Researchers are encouraged to investigate the effects of adaptive digital technologies in terms of learning effectiveness and learning efficiency

Biography

The authors are part of the Faculty of Psychology and Educational Sciences at the KU Leuven, campus Kulak. Dries Debeer is senior researcher and part of the Methodology of the Educational Sciences Research Group. His has expertise in statistical modeling for adaptive learning, prediction and recommendation. Stefanie Vanbecelaere is a PhD student. Her research interests include the effectiveness of adaptive learning of math and reading. Wim Van den Noortgate is full professor at the Faculty of Psychology and Educational Sciences and is affiliated to the Methodology of the Educational Science Research Group. He is a principal investigator of the imec research group itec. His main research focus is on latent variable modeling using multimodal learning analytics, adaptive learning, predication and recommendation and multi-level meta-analysis. 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 investigating cognitive underpinnings of calculation and numerical processes. Fien Depaepe is associate 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 research group itec. Her main research interest deal with the instructional design of technology-enhanced learning environments.

Abstract

During the last decade, many governments and ed-tech companies have demonstrated an increased interest in digital personalized learning, which resulted in a variety of often game-like adaptive learning environments. However, there has been limited attention for the impact of these personalized learning technologies on children's learning efficiency. Does digital personalized learning, like popular claims insist, foster learning in young children? This study attempts to empirically validate the beneficial impact of adaptive learning technology by analyzing log-data from the Number Sense Game (NSG), an educational game that trains early numerical skills. In total, 81 children were randomly assigned to use either an adaptive or a non-adaptive version of the NSG in six sessions in a three-week period. Using a longitudinal item response model children's progress within and across sessions was modeled and compared between the two versions of the game. Regardless of the version of the NSG, children demonstrated progress within and across sessions. However, compared to the non-adaptive NSG, the progress across sessions was stronger in the adaptive NSG. These results provide empirical evidence that adaptive learning environments can improve learning efficiency in young children.

Keywords: adaptivity, log-data, games, early numerical abilities, intervention, learning efficiency

Introduction

An important goal of early education is learners' acquisition of fundamental numerical and reading skills. Teaching these skills to young children requires explicit instruction and varied, intense practice, which should lead to the automatization and fluency that is necessary for obtaining more complex concepts (Reynvoet et al., accepted; Vanbecelaere et al., 2020c). Many children, however, have difficulties acquiring these skills, which might impede their future achievement at school (De Smedt et al., 2013; Sasanguie et al., 2013). Integrating digital educational games in the teaching practice has received much attention as an approach to support the acquisition of numerical and reading skills (Byun & Joung, 2018; Cheung & Slavin, 2013). Indeed, digital educational games have been shown to trigger repeated practice and foster young children's learning (Griffith et al., 2020; O'Rourke et al., 2017; Wouters et al., 2013).

Moreover, compared to traditional paper-and-pen exercises, an advantage of digital educational games is their possibility to automatically present exercises that are adapted to the learner (Plass & Pawar, 2020). Examples of such games in the domain of early numerical and reading skills are "The Number Race" (Wilson et al., 2006), "Math Garden" (Klinkenberg et al., 2011), "Graphogame" (Baker et al., 2017) and the "Number Sense Game" (NSG; Linsen et al., 2015). Within these games, adaptive practice is embedded in an attractive narrative, combined with timely feedback. Furthermore, these games provide instructions audio-visually so that even learners for whom text instructions are too difficult can use the learning environment. In addition, they can provide children with structured sets of exercises enabling repeated practice in a fun and appealing setting (Griffith et al., 2019, 2020).

Often the impact of an adaptive game-based learning environment is evaluated by comparing its use with traditional teaching. However, in such a setting, the added value of adaptivity itself cannot be discerned from the impact of the game-based environment (Clark et al., 2016). Therefore, Vanbecelaere et al. (2020) evaluated the effectiveness of a non-adaptive and adaptive version of the NSG with 78 1st graders. Although they did not find pre-post differences between both versions in standardized tests for numerical competences, they noticed that children playing the adaptive NSG typically required less time to finish the game. One possible explanation for this result is that the adaptive condition learnt more efficiently compared to the non-adaptive condition. An alternative explanation is that the adaptivity allowed students to move through the game faster, without faster learning. The analyses conducted by Vanbecelaere et al. (2020), however, cannot test these two alternative explanations. Therefore, the current study aimed to further explore the results of Vanbecelaere et al. (2020) by modeling learning efficiency using the available log-data from the two versions of the NSG to answer the research question whether adaptivity has a beneficial impact on learning efficiency.

In the following sections we first give a brief overview of the literature on adaptivity in game-based learning. Next, we emphasize the distinction between learning effectiveness and learning efficiency, and explore the potential of log-data to operationalize learning efficiency. In the methods and results sections, we apply a modeling approach to operationalize learning efficiency within and across sessions and investigate the impact of adaptivity in the NSG. Finally, we end with a critical discussion of the obtained results.

Adaptivity in Digital Learning Environments

Plass and Pawar (2020, p. 276) define adaptivity as follows: “the ability of a learning system to diagnose a range of learner variables, and to accommodate a learners’ specific needs by making appropriate adjustments to the learner’s experience with the goal of enhancing learning outcomes”. This definition is broad and there are different ways in which adaptivity has been put into practice. Broadly speaking, implementations of adaptivity in educational games differ regarding three main aspects. First, the learner variables on which the adaptations are based differs across implementations. The adaptivity can, for instance, adapt for learner knowledge, affective states, or level of motivation (Liu et al., 2020; Plass & Pawar, 2020; Vandewaetere et al., 2011). Second, adaptations differ in the specific aspect of the game that is adapted, such as the difficulty level or the presentation of the content, the feedback or the interaction with the player (Liu et al., 2020; Plass & Pawar, 2020). Third, there are different approaches to measure the learner variables which influence the operationalization of the adaptivity itself. Learner variables can be measured before gameplay using pretests or during gameplay based on real-time measurement. The specific operationalizations range from simple threshold rules (e.g., The Number Race) to more advanced techniques to assess children's learning skills, where adaptations are based on expected learner behavior (e.g., Math Garden).

The reasons to make games adaptive are multifaceted. Liu et al. (2020) describe three common ambitions of adaptive educational games. First, they aim to automatically respond to the diverse learners' needs, such as their interests, prior gameplay experience, self-regulated learning skills, or disabilities. Through technology, the process of understanding learners' needs can be facilitated or even automated. Second, because learners' skills are dynamic and constantly changing during gameplay (Thomson et al., 2020), measures recorded at a single timepoint may not paint an accurate picture of the skills. For example, children may start with low gaming abilities but rapidly catch up with their peers. Adaptive games record the learning flow during gameplay and can, therefore, provide exercises that are adapted to the learner's current skill level. Third, and consequently, adaptive technologies can match estimated current skill level with the most appropriate challenge or scaffold. Consequently, scaffolding is realized that matches the zone of proximal development (Vygotsky, 1978). When these ambitions are effectively met, it can be expected that adaptive learning games are beneficial for children's cognitive and non-cognitive learning outcomes (Sampayo-Vargas et al., 2013).

Despite the theoretically expected learning gains, the literature on the effectiveness of adaptive game-based learning presents mixed results. Some studies reported increased learning outcomes due to adaptivity (Sampayo-Vargat et al., 2013, Sandberg, et al., 2014), while others did not find similar effects (Shute et al., 2020; Vanbecelaere et al., 2020b; Van Klaveren et al., 2017). A recent meta-analysis including 12 studies comparing adaptive with non-adaptive practice did not produce a substantial overall effect, when combining effects on learning outcomes, game performance and engagement (Liu et al., 2020). A positive average effect on learning outcomes and a negative effect on in-game performance was observed. The average effect on engagement was also positive, but not significant. Additionally, significant heterogeneity in the effect sizes was found.

We see several issues that could explain the mixed findings on the learning gains of adaptive digital technologies. First, the impact of adaptivity may depend on its actual operationalization. As stated above, where some studies implement adaptivity using simple thresholds, other studies apply more advanced techniques that enable dynamic assessment of the learner's ability and updated estimations of item difficulty (Xie et al., 2019). Second, the instruments that are used to assess the intervention's impact may play a role. Typically, adaptive and non-adaptive environments train the same targeted content and differ only in, for instance, the sequence and/or number of exercises. Thus, large differences in effectiveness are not expected. Frequently, standardized tests are used to evaluate the effectiveness of adaptive games (Vanbecelaere et al., 2020b; Sampayo-Vargas et al., 2013; van Oostendorp et al., 2014; Hooshyar et al., 2018). However, these tests may not be sensitive enough to register differences, because they typically target general competences, whereas many educational games train very specific skills (Reynvoet et al., in press). Finally, many of the studies in the literature do not use an experimental design that allows an unconfounded assessment of the added value of adaptivity. First, non-experimental designs may not completely capture the effect of adaptivity. Second, experimental studies that compare conventional teaching practices with the use of an adaptive digital learning environment (e.g., Faber et al., 2017; Thomson et al., 2020), cannot attribute the found learning gains to adaptivity solely.

Learning Efficiency Vs. Learning Effectiveness

Learning gains can be investigated in multiple ways. For instance, in their framework to evaluate the added value of educational games, All et al. (2015) made a distinction between cognitive, non-cognitive, and efficiency outcomes. The first two outcomes can be related to learning effectiveness and the latter o learning efficiency. Although there are no clear-cut definitions for learning effectiveness and learning efficiency, and despite the fact that they have been mixed up in the literature, there seems to be a new-found consensus on how to interpret learning effectiveness and learning efficiency. In a recent study Molenaar et al. (2019) associate learning effectiveness with whether the learning goals were attained and the learned content could be transferred to other skills and/or the deepness of learning. In contrast, they define learning efficiency as the relation between learners' actions and learning achievements. Learners consider the cost of their actions in light of their learning effectiveness and amount of effort and time invested. Efficient learners yield the highest learning gain in relation to the least effort and time investment (Molenaar et al., 2019).

However, when evaluating learning gains, these gains are most often operationalized as learning effectiveness, in terms of cognitive and non-cognitive outcomes (see Xie et al., 2019 for a systematic review). In contrast, learning efficiency of adaptive game-based learning environments has hardly been investigated. Hence, All et al. (2015) encourage researchers to also consider efficiency outcomes, arguing that when certain skills can be acquired in less time, the investigated educational intervention can be considered effective. There is however, no agreement on how to operationalize or assess learning efficiency, which could be one of the reasons why learning efficiency has not been considered often so far (Vanbecelaere et al., 2020a, van Oostendorp et al., 2014).

One of the first meta-analyses on the efficiency of educational technology reported its potential in substantially saving instruction time (Kulik & Kulik, 1991). Learning efficiency was operationalized as the ratio of the instructional time required for the experimental/computer-based instruction divided by the instructional time required for the conventional/non-computer-based instruction. This operationalization is limited because it can only be used to compare different types of instructions, and because it focuses on the efficiency of the instruction rather than on the learning efficiency itself. Since the meta-analysis of Kulik and Kulik (1991), the computational speed and storage capacities available for digital learning environments have increased greatly. Consequently, learner actions and decisions can be logged and processed with limited costs, providing moment-by-moment learner information. Consequently, log-data can be used to model learning while it happens, thereby presenting alternative and more advanced ways to operationalize learning efficiency. Below we will introduce such an operationalization, which considers the progress in (game) performance, based on moment-by-moment log-data.

Learning While it Happens Via Log-data

There are several reasons why we believe the use of log-data is promising to model learning while it happens, and to assess learning effectiveness and efficiency. First, log-data can provide fine-grained information about students' behavior and performance during the learning process. Thereby, it can give more insights in how (adaptive) games work and interact with learners (Thomson et al. 2020). Second, log-data can be used to model the students' progress over time (i.e., their learning rate) which can be used as a determinant of students' learning efficiency. Third, in comparison to using a wide range of tests that are administered before and after an intervention, analyzing log-data is a non-intrusive way to assess students without requiring a transfer for the tasks in the learning environment to the (standardized) tests (Nebel et al., 2020; Thomson et al., 2020).

Although pre- and post-tests data remain important, the study of Faber et al. (2017) illustrates that log-data can also be used to assess learning effectiveness. Faber et al. (2017) analyzed log-data from an intervention study that compared an online math training with "Snappet" (a digital learning environment) with conventional teaching. They concluded that training with Snappet was more effective. As an alternative approach, in their meta-analysis, Liu et al. (2020) present different game-specific ways to operationalize game performance by means of log-data, such as in-game scores.

Recent studies illustrate that log-data may be even more useful for investigating learning efficiency. For instance, to compare the efficiency of an adaptive and non-adaptive learning environment, Molenaar et al. (2019) assessed learning efficiency by dividing the number of correctly answered problems by the total number of completed problems. In addition, they created moment-by-moment learning curves using Bayesian knowledge tracing, and concluded that the learning curves reflected the efficiency of students' learning regulation over time. Thomson et al. (2020) employed an alternative approach to investigate learning efficiency. In an experimental study over a period of 25 weeks, they assessed progress using growth curves based on log-data. The results yielded that the variation in trajectories obtained by the growth curves could explain the variation in literacy performance better than tests taken at a single timepoint. Finally, Nebel et al. (2020) combined the time learners needed to respond to questions and the number of correct responses to calculate learning efficiency using a very specific formula. They concluded that an adaptive competitive element increased efficiency more than competing against human opponents.

This Study

In this study, we introduce a promising approach to model and assess learning efficiency using log-data. More specifically, we investigate the impact of adaptivity on learning efficiency in an educational game that trains early numerical skills. We analyze the log-data from the study of Vanbecelaere et al. (2020a), which implemented two different versions of the number sense game (NSG; Linsen et al., 2015), to capture moment-by-moment learning. The goal of this study is to investigate whether the adaptive training produced larger learning efficiency compared to a non-adaptive training in a well-controlled setting. Given that it is aimed to validate a given intervention as the cause of cognitive improvements above and beyond any placebo or expectation-related effects, this study can be categorized as an 'efficacy' study (Green et al., 2019). Inspired by the modeling approach proposed by Kadengye et al. (2015), we assess learning progress as a measure of learning efficiency (cf. Appendix B). Our approach has three advantages. First, learning progress can be assessed both within and across game playing sessions. Second, the added value of adaptivity can be added to the model and, hence, directly tested. Finally, by modeling individual item responses differences in item difficulty are taken into account, which is a prerequisite for modeling progress unbiasedly. We expect that the adaptive NSG will lead to stronger progress both within and across sessions, and hence to higher learning efficiency.

Methods

Number Sense Game

The NSG[1] consists of two subgames: (a) a number comparison game (NC) in which two numerosities are presented and the largest has to be selected and (b) a number line estimation game (NLE) in which a numerosity has to be placed on a number line going from one to ten. Both the NC and NLE include domain-specific tasks that train early numerical skills. In the NC, 14 levels are designed to vary in difficulty based on the presented numbers (1-4, 1-9, or 5-18), the display duration (until response or 1500 ms) and the type of stimuli (non-symbolic, symbolic, or mixed notation).

[1] NSG, developed by the GOA consortium led by Lieven Verschaffel – GOA 2012/010

Similarly, in the NLE, there are 18 levels of varying difficulty based on the number of anchor points (9, 1, or, no anchor points) the display duration (until response or 1500 ms), and the type of stimuli (non-symbolic, symbolic, and mixed notation). A detailed description of the NSG can be found in Linsen et al. (2015). Previous intervention studies provided evidence for the NSG training meaning that the NSG training is at least equally beneficial or better in terms of learning as 'business as usual' or active control condition (Maertens et al., 2016; Vanbecelaere et al., 2020b).

Non-Adaptive and Adaptive NSG

Based on the original NSG, which included basic rule-based adaptivity, a non-adaptive and a new adaptive NSG were implemented by Vanbecelaere et al. (2020a). In the non-adaptive NSG, first all the NC and then all the NLE levels were consecutively presented in a fixed order, regardless of the students' performance. Within each level, items were randomly administered. In the adaptive NSG, however, the order in which the levels were presented was adapted to students' performance in the previous levels. The implementation of adaptivity was inspired by the Elo-rating system (Elo, 1978), in which the ability or rating of players is constantly updated after each item response (for a similar approach, see Klinkenberg et al., 2011).

To be more precise, the adaptivity was based on the Rasch model (Rasch, 1980), in which the probability of a correct response is assumed to be a function of the difficulty of the item (τ_i, for item $i = 1, 2, \dots, I$) and the ability of the student (ζ_p for student $p = 1, 2, \dots, P$) (Elo, 1978; Rasch, 1980). However, rather than assuming a constant ability during the game, the ability was allowed to change. Using an updating rule (Elo, 1978) the ability was adjusted after every response, taking into account the accuracy of the response and the expected response given the last ability rating and the difficulty of the item. In contrast with the original Elo-rating, and, for instance, the approach used by Klinkenberg et al. (2011), item difficulties were assumed fixed. That is, the item difficulty parameters were calibrated using data from a previous study and not updated anymore during the current study (Vanbecelaere et al., 2020a).

Figure 2. Screenshots of the comparison game (left) and the NLE game (right)

In the adaptive NSG, first for the NC levels, then for the NLE levels, the next level was chosen, so that its average difficulty was close to the current ability estimate. More precisely, if multiple levels had an average difficulty within a range of 0.2 around the current ability estimate, a random level from this set was selected. Otherwise, the level with the average difficulty that was closest to the current ability level was selected. Like in the non-adaptive NSG, within each level, items were randomly administered. Each student started with the same (low) ability rating in the easiest NC level ($\zeta_{p0,NC}$), and the student's ability was updated after every response throughout the adaptively administered NC levels. After completing the hardest NC level, the adaptive process was repeated, for the NLE levels, again starting from the easiest NLE level and from the same (low) ability rating ($\zeta_{p0,NLE}$) for all students.

Procedure

First grade students (*N* = 84, 39 girls) from five classes in three different elementary schools took part in this study. During three weeks, students participated in six 30-minute sessions (two sessions per week) in which they played the NSG in class. Most students finished the NSG before the last session. Students were randomly and blindly assigned to either the adaptive (45 students) or non-adaptive NSG (39 students). After excluding three students who were absent during more than one session, the results are based on 81 students: 44 in the adaptive and 37 in the non-adaptive NSG.

In each session, a researcher went to the classroom to distribute tablets and headphones, and to support the students during the game sessions. The participants were administered a set of cognitive and non-cognitive tests before the first session, directly after the last session, and two weeks after the last session. In this manuscript, however, the focus is on the log-data and moment-by-moment learning, the data from the tests are not discussed. For a complete description and analysis of the administered tests, see Appendix A and Vanbecelaere et al. (2020a).

Modeling Learning Efficiency

Item-wise accuracy data extracted from the log-data were analyzed using a modeling approach based on Kadengye et al. (2015). This approach uses a longitudinal generalized mixed modeling framework to model moment-by-moment learning. The modeling approach has been previously used to model learning in, for instance, an online learning environment with statistics exercises (Kadengye, et al., 2015) and in Massive Open Online Courses (MOOCs; Abbakumov et al., 2019). We briefly describe the most relevant parameters of the modeling approach here, a detailed discussion is provided in Appendix B.

The modeling approach models the probability of a correct response while considering (a) differences in difficulty between the items, (b) differences in ability between the students, and (c) learning both within and across sessions. To model learning within sessions, the model includes a parameter α_1 that can be interpreted as the average learning rate (or learning efficiency) within sessions (i.e., how much does a student on average improve within a session?). Likewise, learning across sessions is modeled by including a parameter α_2 that can be interpreted as the average learning rate over sessions (i.e., how much does a student on average improve from one session to the next sessions?). Individual differences across students in the learning rates within and over sessions are also included in the model. Finally, the underlying abilities at the end of the game-playing sessions are also modeled. The complete formulation of the model is given in Appendix B.

Hypothesis Testing

The main aim of the analysis is to investigate whether adaptivity in the NSG positively affects learning efficiency. In addition, the impact of adaptivity on the ability at the end of the last session is investigated. The following three hypotheses are of interest:

1. Adaptivity increases the learning rate within a session: The learning rate within a session is higher in the adaptive NSG than in the non-adaptive NSG.
2. Adaptivity increases the learning rate across sessions: The learning rate over sessions is higher in the adaptive NSG than in the non-adaptive NSG.
3. Adaptivity increases the final ability: The ability at the last response of the last session is higher in the adaptive NSG than in the non-adaptive NSG.

The adaptivity hypotheses were tested using three one-sided Wald tests. Positive parameter estimates correspond to the hypotheses that adaptivity increases learning efficiency within and over sessions, and increases the final ability, respectively.

Results

Descriptive Statistics

Table 1 presents the average number of items students responded to and the average proportion correct responses for both versions of the NSG. In the adaptive NSG, students generally responded to less items compared to the non-adaptive NSG. The biggest difference is found in the NC game, where students playing the adaptive NSG responded to more than 55 percent less items than students in the non-adaptive NSG. The differences in the average proportion of correct responses is less prominent. Although differences are small, the average proportion correct responses in the NC game was higher in the non-adaptive NSG, whereas for NLE items, the adaptive NSG leads to the highest proportion correct responses.

Table 1

Number of Administered Items and Proportion of Correct Responses in the Non-Adaptive and the Adaptive NSG

Game Version	N	Number Items			Proportion Correct		
		NC	NLE	Total	NC	NLE	Total
Non-adaptive	37	344.9	399.7	684.6	0.899	0.675	0.788
Adaptive	44	149.0	263.1	412.0	0.868	0.711	0.762

Figure 1 presents density approximations for the number of items that a student responded to

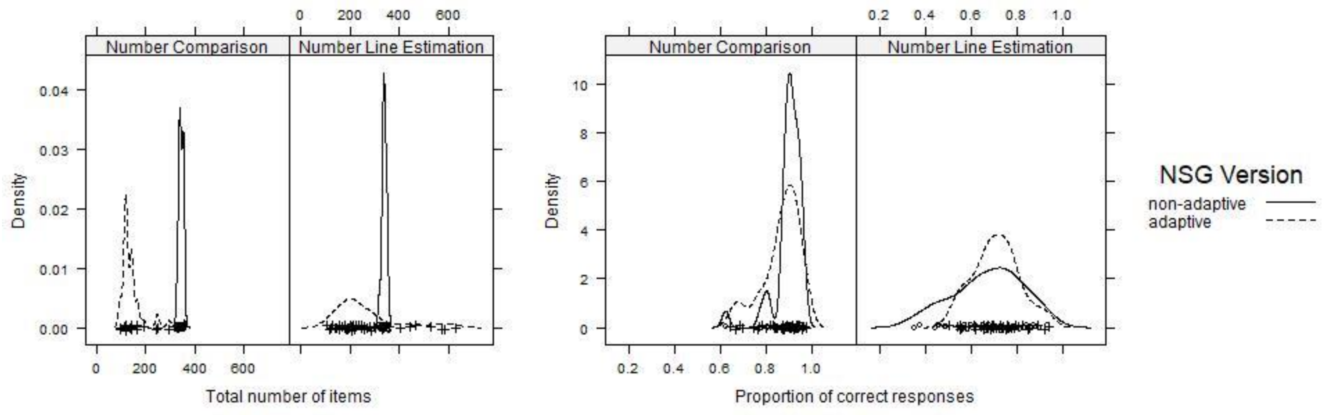

Figure 1. Density approximation for the number of items and the proportion correct for the non-adaptive and the adaptive NSG. Separate densities are drawn for the NC and the NLE levels.

These descriptives show that students in the adaptive condition were administered less than half of the NC items. Indeed, half of the NC levels were played by only 20% or less of the students assigned to the adaptive condition. More than 85% of the students in the adaptive condition only played NC levels during the first session. Furthermore, there was little variation in the timepoints at which students were administered the NC levels, even in the adaptive condition. Because variation across students in the timepoints at which they respond to specific items (or levels) is a prerequisite for modeling learning in our approach[2], we chose to exclude the NC responses from the analysis. Hence, the results in the following section are only based on the NLE items. For completeness, the appendices also provide the results based on the complete data, with similar findings.

Hypothesis Testing

Because the analysis was based on NLE items only, the ability (at the last response of the last session) and the learning rate within and across sessions can be interpreted as the NLE ability and the NLE learning rate within and across sessions. The results provide evidence for a positive NLE learning rate within sessions in the non-adaptive (estimated value of .41) as well as in the adaptive condition (estimated value of .84). We found limited evidence for a higher NLE within session learning efficiency in the adaptive compared to the non-adaptive condition ($z = 1.82$, $p = .034$). Further a positive NLE across sessions learning rate was found in the non-adaptive condition (estimated value of .35). Yet there was evidence that the NLE learning rate was even stronger in the adaptive condition ($z = 2.27$, $p = .011$) with an estimated value of .86. In contrast, no evidence was found for differences in the NLE ability at the last response of the last session between the non-adaptive and the adaptive condition ($z = 1.40$, $p = .081$).

[2] The levels in the non-adaptive NSG were ordered according to increasing difficulty as evaluated by educational experts. Differences in the order of the levels in the adaptive NSG are required to disentangle the effect of learning and the effect of level difficulty.

Discussion

The present study was designed to investigate whether adaptivity in the NSG, an educational game that trains early numerical skills, has a beneficial impact on learning efficiency, by analyzing moment-by-moment log-data. In an experimental design with 81 first grade students, we tested whether the learning efficiency (within and across game-playing sessions) in an adaptive NSG was higher than in a non-adaptive NSG. Log-data were analyzed using a modeling approach that accounts for differences in the difficulty of the items and differences in the ability of the students when assessing students' progress.

In our analysis using only the NLE items, we found that students made progress both in the non-adaptive and adaptive condition, but that the adaptive NSG stimulated learning more. The implemented adaptivity increased learning efficiency across, but not within a game-playing session. The higher learning efficiency in the adaptive NSG, however, did not lead to higher estimated abilities at the last response of the last session. This observation corresponds to the findings of Vanbecelaere et al. (2020a) based on a pre-post test analysis, and provides evidence that, though not improving learning effectiveness, adaptivity in educational games can foster learning efficiency.

In the remainder of this discussion we will first relate our results to previous findings. Then we compare our operationalization of learning efficiency with other operationalizations. Subsequently we discuss the practical implications of our findings. After critically discussing the limitations of our study, we give directions for possible extensions and future research.

Previously, mixed results were found regarding the learning gains due to adaptivity in educational games (Liu et al., 2020). Most studies, however, operationalized learning gains in terms of learning effectiveness and not in terms of learning efficiency. Furthermore, often pre-post-test differences were assessed, which assume that the skills trained in the digital environment transfer to standardized tests. In contrast, this study analyzed in-game log-data to assess learning while it happens and focused on learning efficiency, which relates to the time required to master a skill (Molenaar et al., 2019). Although it is not clear whether a focus on learning efficiency (operationalized using log-data) could lead to more consistent findings, we would encourage researchers to also consider learning efficiency outcomes when assessing the impact of adaptivity in digital learning environments (cf., All et al., 2015).

Assessing learning efficiency using in-game log-data is not new. Yet the way we operationalized learning efficiency, differs from previous studies (Molenaar et al., 2019; Nebel et al., 2020). First, we modeled moment-by-moment learning at the more fine-grained level of the responses to items while, for example, Nebel et al. (2020) assessed learning efficiency based on the mean number of correct answers and the average time students needed to complete a task. Second, we assessed learning efficiency on a continuous dimension that is directly related to the progress made over time. In contrast, Molenaar et al. (2019) used Bayesian Knowledge Tracing and visual inspection to discern four types of moment-by-moment learning curves (Baker et al., 2013), and related these curve types to learning efficiency. Although their methodology differs, item-level data also lies at the basis of their approaches. Therefore, it would be interesting to relate our two operationalization of learning efficiency with the approach of Molenaar et al. (2019), both theoretically and empirically (i.e., do the different types of moment-by-moment curves correspond to differences in the assessed learning rate?).

Implications for learning practice

Because we found evidence for a beneficial effect of adaptivity on learning efficiency in NLE items, further development and implementation of adaptivity in digital learning environments should be encouraged. However, even without an impact on learning gains we would still advocate for the use of adaptivity. An adaptive system differentiates based on learner behavior. For instance, in our study, the variation in the number of NC and especially NLE items indicates that the adaptive NSG adapted the learning pace to individual learners (cf. Figure 1). As a result, faster learners were finished after three to four sessions, while slow learners were given more time to master the skills. The non-adaptive NSG did not allow this diversification: all learners had to follow the same prescribed learning path. We believe that in learning, it is important to considerer individual differences and act upon those to stimulate the learning process for all learners (Dowker, 2017). Adaptivity can assist in reaching this goal.

The potential differentiation in adaptive learning environments also illustrates the need for proper teacher dashboards (Molenaar et al., 2019). Teachers should be able to get immediate feedback about the progress learners are making. Then, when slower learners are identified, teachers can decide whether or not to intervene, or give more specialized (personalized) learning activities. Likewise, when faster learners are detected, teachers can challenge them with new learning activities, or motivate them to further train the acquired skills to make sure they get enough practice. With efficient feedback from dashboards, adaptivity within a learning environment can lead to further personalization by teachers.

Limitations

This study focused on the interaction between the digital learning environment and the learner. That is, although the learning sessions took place in the classroom, involvement of the teacher (or researcher) was artificially prevented to standardize the intervention. However, it is clear that the role of the teacher should not be underestimated, even when the learning takes place in a digital environment. So far, the mediating or moderating role of the teacher in the learning environment has received little attention (Molenaar et al., 2019). Teacher actions may, for instance, reduce the learning efficiency differences between learners due to adaptive and non-adaptive learning environments. Alternatively, when informed by efficient dashboards, teachers could even boost the learning efficiency gains due to adaptivity.

Like many other studies investigating the impact of adaptivity on learning gains, we focused on a specific learning environment developed for a specific age-group, training a specific skill, combined with a specific implementation of adaptivity. Consequently, we should be careful with generalizations to other learning environments, skills, age-groups or adaptivity implementations. Nevertheless, we expect similar findings for other drill-and-practice games focusing on early skills that require intensive and repeated practice such as early reading skills (Thomson et al., 2020). Of course, future research should test these claims.

From a methodological perspective, ideally all levels and items should have been randomly administered in the non-adaptive NSG. That is, differences in the order and time at which items are administered are a prerequisite for capturing learning. Yet from a didactical perspective complete randomness would have been highly inappropriate, because this could have resulted in the presentation of the most difficult items at the start of the first session. These items would not foster learning in most students, and in addition, the responses to these items would have been uninformative with respect to modeling progress or ability. Therefore, the level order that was supported by the educational experts involved in developing the NSG (Linsen et al., 2015) was preserved in the non-adaptive version. Consequently, we noticed that there was little variation in the timepoints at which students administered the NC levels, even in the adaptive NSG, which made us exclude responses to the NC items from the analysis. For the sake of completeness, Appendix C and D present analyses on the complete data with similar conclusions.

In the adaptive NSG, levels rather than items were selected adaptively. This could explain why we only found evidence for an increased learning rate across game-playing sessions. Although students typically played 5 to 8 levels within a session, this may not have been enough for the adaptivity to have a clear effect within sessions.

Extensions

Our results provide evidence that adaptivity in digital learning environments fosters learning efficiency. Future research could investigate which mechanisms induce this effect. Molenaar et al. (2019), for instance, explained the beneficial impact of adaptivity using self-regulated learning theory. They argued that adaptivity takes over monitoring, planning and evaluating tasks that are required for self-regulated learning but which are often difficult for young children. As such, they claim, adaptivity supports self-regulated learning, thereby increasing learning efficiency. In our study, however, the control and regulation of the learning pace was already present in the non-adaptive NSG (Linsen, et al. 2015), which suggests that there could be other mechanisms that mediate or moderate the adaptivity induced learning efficiency gains. Possible mechanisms could be related to motivation (Faber et al., 2017; Sampayo-Vargas et al., 2013) or scaffolding learning in the zone of proximal development (Vygotsky, 1978). For instance, future research could investigate the role of scaffolding learning in the zone of proximal development by comparing two adaptive versions of the same game-based learning environment, where one version is based on scaffolding principles, and the other is not. As an example, the original NSG could be compared the adaptive NSG version as used in this study. Both versions have adaptivity implementations, but only the version used in this study is based on the principle of scaffolding learning.

In our study, we investigated the average increase in learning efficiency due to adaptivity. However, it can be expected that there are individual differences in the impact of adaptivity. For instance, the impact of adaptivity could be related to learner characteristics, which could be identified either before (e.g., prior knowledge, technology skills, individual risk aversion) or during the learning activity (e.g., motivation, attention) (Vandewaetere et al., 2011). For instance, recently learning characteristics assessed before the learning activity have been used to alleviate the cold start problem that is common to many adaptive systems (Park et al., 2019). Moreover, if future research indicates that some learners benefit more from one type of adaptivity, while other learners benefit from another type of adaptivity, the adaptivity mechanism itself could be adapted based on learner characteristics.

In a next step, field data could validate our findings in real settings. The data collection of Vanbecelaere et al. (2020a) took place in a largely controlled setting: training sessions were strictly timed, the researcher conducted all training sessions with limited influence of the teacher. This approach was necessary to validate the adaptive intervention as the cause of the cognitive improvements above and beyond the non-adaptive training. Following recent recommendations of Green et al. (2020), future research could assess whether this adaptive training also produces a positive impact in real world situations.

To conclude, learning efficiency was operationalized by modeling children's progress within and across sessions, using the log-data collected during gameplay. We found that training early numerical skills with an adaptive digital educational game increased learning efficiency across game-playing sessions when compared a non-adaptive version of the same game. In addition, our approach illustrates how log-data can be used to asses moment-by-moment learning, and to evaluate the effect of adaptivity in terms of learning efficiency.

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, BLCC and Sensotec. The Number Sense Game (NSG) used in this study was developed by the GOA consortium led by Lieven Verschaffel – GOA 2012/010. The LEAPS project was carried out within the imec's Smart Education research programme, co-financed by imec and received project support from Flanders Agency of Innovation & Entrepreneurship. Parts of this work were done within the i-Learn project, funded by Flanders Agency of Innovation & Entrepreneurship (AH.2019.051).

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

Prior to the study, approval of the social and societal ethics committee (SMEC) 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 interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Abbakumov, D., Desmet, P., & Van den Noortgate, W. (2019). Measuring growth in students' proficiency in MOOCs: Two component dynamic extensions for the Rasch model. *Behavior research methods*, *51*(1), 332-341.

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

Baker, R. S., Hershkovitz, A., Rossi, L. M., Goldstein, A. B., & Gowda, S. M. (2013). Predicting robust learning with the visual form of the moment-by-moment learning curve. *The Journal of the Learning Sciences, 22*(4), 639–666. https://doi.org/10.1080/10508406.2013.836653.

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01.

Byun, J., & Joung, E. (2018). Digital game-based learning for K–12 mathematics education: A meta-analysis. *School Science and Mathematics, 118*(3-4), 113-126.

Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education*, *105*, 14-30.

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.

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48-55.

Dowker, A. (2017). Interventions for primary school children with difficulties in mathematics. In *Advances in child development and behavior* (Vol. 53, pp. 255-287). JAI.

Elo, A. E. (1978). *The Rating of chessplayers, Past and Present.* New York: Arco Publishing, Inc.

Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & education, 106,* 83-96.

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., ... & Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, *3*(1), 2-29.

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., & Bagner, D. M. (2020). Apps as learning tools: a systematic review. *Pediatrics, 145*(1).

Griffith, S. F., Hanson, K. G., Rolon-Arroyo, B., & Arnold, D. H. (2019). Promoting early achievement in low-income preschoolers in the United States with educational apps. *Journal of Children and Media*, *13*(3), 328-344.

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.

Kadengye, D. T., Ceulemans, E., & Van Den Noortgate, W. (2015). Modeling growth in electronic learning environments using a longitudinal random item response model. *The Journal of Experimental Education, 83(2)*, 175-202.

Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. (2011). Computer adaptive practice of math 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.

Liu, Z., Moon, J., Kim, B., & Dai, C. P. (2020). Integrating adaptivity in educational games: a combined bibliometric analysis and meta-analysis review. *Educational Technology Research and Development, 68*, 1931–1959. https://doi.org/10.1007/s11423-020-09791-4

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: Does it matter?. *Learning and Instruction*, *46*, 1-11.

Molenaar, I., Horvers, A., & Baker, R. S. (2019). What can Moment-by-Moment learning curves tell about students' self-regulated learning? *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2019.05.003

Noroozi, O., Pijeira-Díaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirschner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. Education and Information Technologies, 1-49.

O'Rourke, J., Main, S., & Hill, S. M. (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.

Park, J. Y., Joo, S. H., Cornillie, F., van der Maas, H. L., & Van den Noortgate, W. (2019). An explanatory item response theory method for alleviating the cold-start problem in adaptive learning environments. *Behavior Research Methods, 51*(2), 895-909.

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275-300.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment Tests*. Chicago: The University of Chicago Press.

Reynvoet, B., Vanbecelaere, S., Sasanguie, D., & Depaepe, F. (in press). Intervention studies in math: A discussion. In W. Fias & A. Henik (Eds.), *Learning and education in numerical cognition*. Elsevier.

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 & Education, 69*, 452-462.

Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers & Education*, *76*, 119-130.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*, 418-431.

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C. P., ... & Sun, C. (2020). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12473

Thomson, J. M., Foldnes, N., Uppstad, P. H., Njå, M., Solheim, O. J., & Lundetræ, K. (2020). Can children's instructional gameplay activity be used as a predictive indicator of reading skills? *Learning and Instruction, 68*, 101348.

Vanbecelaere, S., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020a). The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, noncognitive and efficiency outcomes. *British Journal of Educational Technology*.

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020b). 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. (2020c). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning, 36*(4), 502-513.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*(1), 118-130.

Van Klaveren, C., Vonk, S., & Cornelisz, I. (2017). The effect of adaptive versus static practicing on student learning-evidence from a randomized field experiment. *Economics of Education Review*, *58*, 175-187.

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 Trans. *Serious Games, 1*(2), e5.

Vygotsky, L. S. (1978). *Mind in society. The development of higher cognitive processes.* Cambridge, MA: Harvard University Press.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of" The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*,*2*(1), 19.

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.

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education, 140*, 103599.