Computers & Education 101 (2016) 168–178

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Discovering the laws of physics with a serious game in kindergarten

Joep van der Graaf*, Eliane Segers, Ludo Verhoeven

Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 20 December 2015
Received in revised form 21 June 2016
Accepted 26 June 2016
Available online 27 June 2016

Keywords:
Scientific thinking
Serious games
Interactive learning environments
Individual differences
Kindergarten

ABSTRACT

Serious games have unique strengths that can be used to augment science education. For the current study, we developed and investigated a serious game to assess kindergartners' discovery of the laws of physics in the so-called Hippo app. The participants were 71 children, aged 5 years and 5 months on average. The app consisted of three game-plays: slides, seesaw, and pendulum. Children were asked to set variables (such as the steepness of the slide) correctly in order to provide a hungry hippo with a drink or some food. Children's gaming behavior was assessed via exploration and efficiency scores, and next related to executive control, nonverbal reasoning, and vocabulary. Exploration was defined as the number of actions corrected for the total playing time, efficiency as the number of correctly solved problems corrected for the total number of attempts. The results revealed that efficiency and exploration scores did not correlate significantly, indicating two distinct types of gaming behaviors. Both types were associated with attentional control. Mediation analysis showed that the relation between exploration and attentional control was mediated by vocabulary, while the relation between efficiency and attentional control was mediated by nonverbal reasoning. To conclude, kindergartners' efficiency and exploration can be seen as independent game behaviors; both demanding attentional control, but mediated by verbal skills in the case of exploration and by nonverbal reasoning in the case of efficiency.

1. Introduction

Scientific thinking has become more relevant in recent years. Many countries aim to stimulate scientific thinking via science education in order train future researchers, but also to make children think critically (Osborne, 2013). Serious games may very well be a supportive tool in this respect (Hamari, Koivisto, & Sarsa, 2014; Morris, Croker, Zimmerman, Gill, & Romig, 2013), and they also allow the assessment of *how* children play them. As a case in point, a physics game may help identify exactly how young children discover scientific concepts. In kindergarten, there is already some knowledge of physics (Inhelder & Piaget, 1958), but little is known about how children gain this knowledge, and how individual differences can be explained. In a game-like assessment, children would require explorative and efficient behavior. Exploration is about where to look, while efficiency is about understanding what is seen (Klahr & Dunbar, 1988). Both exploration and efficiency have been

* Corresponding author. Behavioural Science Institute, Radboud University, Montessorilaan 3, Room A05.16, P.O.Box 9104, 6500 HE, Nijmegen, The Netherlands.

E-mail address: j.vandergraaf@pwo.ru.nl (J. van der Graaf).

http://dx.doi.org/10.1016/j.compedu.2016.06.006

shown to foster scientific thinking, but which factors underlie these effects is largely unknown, especially in young children (see Legare, 2014 for a review). To identify the basis of exploration and efficiency, both cognitive (executive control, reasoning ability), and linguistic (vocabulary) factors may play a role (Wagensveld, Kleemans, Segers, & Verhoeven, 2015). In the present study, we therefore assessed individual differences in exploration and efficiency during physics gaming in kindergarten.

1.1. Serious games to enhance scientific thinking

To participate in our contemporary knowledge-based society, a broad set of skills and abilities are required, such as scientific thinking (Dede, 2010). Scientific thinking consists of scientific reasoning, i.e. domain-general skills, and scientific knowledge, i.e. domain-specific knowledge (Klahr, Zimmerman, & Jirout, 2011). Scientific reasoning is the intentional seeking of knowledge using the scientific method (Kuhn, 2004). The three components of scientific reasoning are: hypothesis generation, designing and conducting experiments, and evaluating the evidence (Klahr, 2000). These three components allow people to evaluate the validity of science-related claims, and to generate new knowledge (Fischer et al., 2014), which helps them to become critical and autonomic thinkers. To support the development of these skills, scientific activities are being undertaken in schools (e.g., Lorch, Lorch, Freer, Dunlap, & Hodell, 2014). Children can acquire scientific knowledge about specific domains via these scientific activities. An example is learning why objects float or sink (Hardy, Jonen, Möller, & Stern, 2006). Children can acquire this domain-specific knowledge via exploration and efficiency (Klahr & Dunbar, 1988). Via exploration, children can identify what the factors are that determine the effect, such as whether a boat will float. Via efficient behavior, children can generate a desired effect, such as designing a boat that will float. These types of behavior underlie scientific thinking (Legare, 2014).

A recent approach to scientific thinking in education is the use of serious games. Games provide a rich learning environment and allow children to select their own choices, which strengthens their motivation (Morris et al., 2013). The richness of serious games motivates children to process multiple, possibly interacting variables, which makes them especially suitable for science education (Morris et al., 2013). Just as rule discovery and induction are inherent to scientific thinking (Dunbar & Klahr, 2012), so are they to gaming (Greenfield et al., 1994). The richness of serious games also provides an opportunity for exploration, as there is more vigorous and prolonged exploration of stimuli that can be described as complex (Berlyne, 1966). Another advantage of serious games is that they can be regarded as ongoing assessments, providing players with the opportunity to show what they have learned (Morris et al., 2013). Learning is not linear, but multiple strategies coexist to solve a problem, which is affected by experience (Siegler, 2000), which seems to be the case in serious games as well. Therefore, serious games seem an ideal tool to assess children's explorative behavior and problem-solving skills within science education.

1.2. Discovering the laws of physics

In kindergarten, children already show the ability to think scientifically, which is about scientific reasoning and scientific knowledge. Kindergartners can reason scientifically, although they have great difficulty with hypothesis generation (Piekny & Maehler, 2013). However, many kindergartners are able to set variables correctly in order to investigate one of them. While performance decreased when the number of variables increased, many kindergartners turned out to be able to design multivariable experiments in a study by Van der Graaf, Segers, and Verhoeven (2015). It has also been shown that kindergartners can evaluate various types of evidence (Piekny & Maehler, 2013) which indicates that they have the ability to obtain domain-specific scientific knowledge themselves. Indeed, young children already show considerable conceptual change in the domain of physics, with or even without instruction (Vosniadou, 2002). In classic experiments, it has been shown that many young children understand some physics, such as the effect of the angle of a slope on how far a ball will roll (Inhelder & Piaget, 1958). They also understand how weight affects the balance of a balance beam, although they find it difficult to identify distance to the fulcrum as a variable (Inhelder & Piaget, 1958; Siegler, 1976). Using a pendulum, young children can find correspondences between the length of the string and the frequency of the oscillations, but they do not grasp the mechanism of the pendulum completely, because they fail to distinguish between their own actions and the motion of the pendulum itself (Inhelder & Piaget, 1958).

In discovering the laws of physics, exploration is needed, as well as a certain level of efficiency in problem solving. Exploration is needed to identify variables and their effects, while efficiency is needed to solve problems in a goal-directed manner. These skills have been described by Klahr and Dunbar (1988) within scientific thinking as: "The successful scientist, like the successful explorer, must master two related skills: knowing where to look and understanding what is seen" (p. 2). The first skill, exploration, involves the design of experimental and observational methods, and the second skill, efficiency, involves the formation and evaluation of theory (Klahr & Dunbar, 1988). Exploration helps children in making their prior beliefs more explicit, and there is more exploration when evidence contradicts their prior belief than when evidence confirms their prior belief (Bonawitz, Van Schijndel, Friel, & Schulz, 2012). It has also been suggested that exploration is especially relevant for young children, because of its relation to curiosity (Cecil, Gray, Thornburg, & Ispa, 1985).

Efficiency can be defined as the ability to solve as much problems in as few possible attempts. Within scientific thinking, efficient behavior would imply that one has understood the underlying mechanisms via which a certain scientific phenomenon works, because one can use it more or less instantly to solve a problem. It has been shown that when children are provided with an explanation, they can recognize information as evidence by incorporating it into a causal framework

(Koslowski, Marasia, Chelenza, & Dublin, 2008). Without this prior knowledge, children may be unable to consider new or alternative explanations.

1.3. Cognitive factors in exploration and efficiency

Both exploration and efficiency have been shown to foster causal reasoning, but factors that underlie the individual differences are largely unknown, especially for young children (Legare, 2014). With regard to individual differences in exploration and efficiency during a physics game, executive control may be a key factor. Young children use executive control during scientific thinking for the rejection of intuitively derived misconceptions (Kwon & Lawson, 2000). Executive control has been described as consisting of an action and an attentional component (Diamond, 2013). Action control (also called behavioral control or response inhibition) is the ability to restrain one's normal or habitual response, which allows children to stop and think during task performance in order to perform in a goal-directed manner, whereas attentional control is the ability to focus on the task, or on aspects of it, allowing children to keep track of information and to adjust their responses accordingly (Cartwright, 2012). Action control may be positively related to game exploration, because the control of one's own actions allows one to guide their own exploration, instead of randomly interacting with the game. The inhibition of automatic responses also allows one to think before acting, which would indicate a positive relation between action control and game efficiency, because one can think of a correct response. The other component of executive control, attentional control can help in scientific thinking by inhibiting prior beliefs (Kuhn & Franklin, 2006). When prior beliefs exist about the rules of the game, such as the effect of certain variables, this might limit exploration, as the child might think it already knows how the game works. Attentional control may therefore be positively related to game exploration. With regard to efficiency, when a counterintuitive response is required, the inhibition of prior beliefs helps in creating the correct response. Attentional control also enhances the formation of problem-solving strategies, formulating strategies and more (Fernyhough & Fradley, 2005). For efficient scientific discovery, problem-solving strategies are required (Dunbar & Klahr, 2012). Attentional control is likely to enhance game efficiency via the formulation and use of strategies.

The relations of executive control with game exploration and efficiency may also be indirect, since executive functions have been found to be related to skill acquisition (Lefevre et al., 2013) and are therefore at the basis for young children's learning (Diamond, 2013). The indirect effects might be via verbal (i.e. vocabulary) and nonverbal (i.e. reasoning) skills. Correct experimentation has been linked to vocabulary and syllogistic reasoning (Wagensveld, Segers, Kleemans, & Verhoeven, 2015). Vocabulary can help children to understand scientific principles. Academic vocabulary allows children to write and talk about science (Snow, 2010) whereas non-academic vocabulary supports the learning of new concepts (Werker, Fennell, Corcoran, & Stager, 2002). In the case of discovering physics with a game, the variables are the new concepts to be learned. Before the variables can be explored, they should be identified and encoded. In the visual domain, Dunbar and Klahr (2012) proposed nonverbal reasoning to be involved, because it allows one to discover a rule that explains a series of events, or extrapolate from a rule to formulate theories and yet-to-be observed phenomena. Indeed, nonverbal reasoning was related to correct performance on experimentation in kindergarten (Van der Graaf et al., 2015). Efficient behavior when playing serious games may therefore be related to reasoning abilities. Because vocabulary relates to (learning of) new concepts (Werker et al., 2002) and these can aid exploration, game exploration may rely more on vocabulary. On the other hand, nonverbal reasoning has been shown to relate to performance in science (Van der Graaf et al., 2015), and thus game efficiency may rely more on nonverbal reasoning.

1.4. The present study

To summarize, young children show abilities to participate in scientific discourse and they already have some understanding of physics. However, little is known about how children discover the laws of physics. In the present study, we used a serious game to assess this and investigated the cognitive factors involved in game exploration and game efficiency. The game provided children with an informal environment in which they can explore and have opportunities for learning (cf. Legare, 2014). The game allowed children to discover the laws of physics in three problem-based environments inspired by the work of Inhelder and Piaget (1958): a slide, a seesaw (balance beam), and a pendulum. We assessed children's exploration, measured as actions per second, and efficiency, measured as proportion correct per attempt for each of these environments. We had the following research questions:

1. To what extent do children show comparable exploration and efficiency scores on the game-plays of slides (slopes), seesaw (balance beam), and pendulum, and on the levels of the game-plays?
2. How does the individual variation in exploration and efficiency relate to executive control, reasoning ability, and vocabulary?

With respect to the first question, we expected no differences for the exploration scores. The exploration score was a measure of the interaction with the variables. Because each game had the same amount of variables, there was no reason to assume that the game-play would affect the exploration score. For the efficiency scores, however, we hypothesized in line with the studies by Inhelder and Piaget (1958) that the children would be more efficient solving problems with the slides,

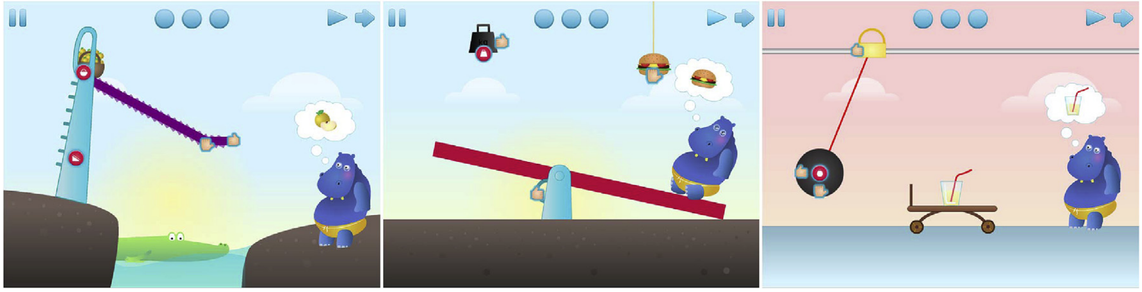

Fig. 1. Screenshots of the game-plays of the Hippo app: the Slides, the Seesaw, and the Pendulum. For the Seesaw in the current screenshot, the child can adjust four variables: the heaviness of the weight, the placement of the weight, the placement of the fulcrum, and the height of the hamburger.

followed by the seesaw, and that they would be least efficient in solving problems with the pendulum. It has been shown when the number of variables increases, performance drops, for example when incorporating weight, besides distance, in kindergartner's prediction of which side of the balance beam would go down (Siegler, 1976). It was therefore expected that performance increased when the number of variables decreased as they moved to a higher game level. With regard to the second question, we assumed children's exploration and efficiency scores to be unrelated, and expected that both scores would be predicted by executive control. In addition, we predicted that the relation between executive control and exploration scores would be mediated by verbal abilities and the relation between executive control and efficiency scores by reasoning abilities.

2. Methods

2.1. Participants

The participants were 75 children from two elementary schools; 41 girls and 34 boys. The children were in the first or second year of kindergarten, which is a two year program in the Netherlands before formal education starts. The average age of the children was 5 years and 5 months (range 4 years and 2 months to 6 years and 7 months). The children were from two schools. One school participated in a previous study (Van der Graaf et al., 2015). We asked their parents for active consent for the present study. The parents of 24 out of 46 children returned the consent letter with a positive answer. The other parents did not return the consent letter, although we had asked for an active "no" if they did not agree. It can be inferred that many just forgot to return the letter. There was no difference in the ratio of boys and girls, $t(44) < 0.01$, $p > 0.999$, age, $t(43) = 0.41$, $p = 0.684$, scores on attentional and action control, $t(41) = 1.30$, $p = 0.202$ and $t(41) = 0.10$, $p = 0.918$, respectively, and scores on nonverbal reasoning and vocabulary, $t(44) = 0.66$, $p = 0.466$ and $t(44) = 1.21$, $p = 0.636$, respectively, between children who were allowed to participate and those who were not. The parents of 65 children from the other school returned the consent letter with a positive consent. Because 14 children of this school did not finish all tests, due to time restrictions, the sample size of the Hippo app was 75 and 71 children performed all tasks.

2.2. Materials

2.2.1. Hippo app

The Hippo app (designed by the authors in cooperation with Qlvr®, for the purpose of the current study[1]) is a serious game, which measures children's discovery of the laws of physics. The app consists of three different game-plays, which address different physics topics: the Slides, the Seesaw, and the Pendulum (see Fig. 1). The goal of each game-play was to provide the hippo with some food or with a drink. This could be done by setting the variables so that the food or drink would end up in the mouth of the hippo. Before the games were played, children were introduced to the use of a tablet. The experimenter showed how the tablet worked and how to set the variables in the games. Each game consisted of a practice round and four levels. The number of variables, that the children could set, decreased with level, from four variables to one variable. All variables had five possible settings.

The goal of the Slides was to slide a basket of apples into the hippo's mouth. The variables of the Slides were the length of the slope, the steepness of the slope, the weight of the basket of apples, and the surface texture of the slope. The goal of the Seesaw was to drop a weight on one side of the seesaw to launch the hippo, that stood on the other side of the seesaw, into the air to allow the hippo near a hamburger, which he could then eat. The variables of the Seesaw were the heaviness of the weight, the placement of the weight relative to the fulcrum of the seesaw, the placement of the fulcrum, and the height at which the sandwich was hanging. The goal of the Pendulum was to provide the hippo with a glass of lemonade. This could be done by letting a weight on a string bump into a cart with a glass of lemonade on it so that the cart would ride near the hippo.

[1] The authors declare to have no financial interest in the Hippo app.

The variables of the Pendulum were the heaviness of the weight, the placement of the string on the ceiling, the height of which the weight could swing, and the placement of the weight compared to the cart (i.e., the horizontal difference in distance between the weight on the string and the cart with a glass of lemonade on it).

2.2.1.1. Gaming. Whenever the hippo was provided with the food or the drink, the attempt was scored as correct. The app provided feedback on whether the attempt was correct or not. There were three possible attempts per level. When an attempt was incorrect the experimenter encouraged the child to try again. Whenever the attempt was correct or whenever three attempts were incorrect, the children proceeded to the next level. There was no time limit. The maximum number of correct was one per level, four per game-play, and twelve in total. The gaming behavior that was scored was the correct score, the number of attempts, the amount of actions, and the playtime. Whenever the child adjusted a variable, it was counted as an action. The gaming measures were all recorded in a log file. The log files were used for the analyses.

2.2.1.2. Game exploration. An exploration score was calculated by dividing the total amount of actions ($M = 70.73$, $SD = 27.12$) by the total playtime in seconds ($M = 787.03$, $SD = 250.84$).

2.2.1.3. Game efficiency. An efficiency score was calculated by dividing the total correct score ($M = 6.35$, $SD = 1.98$) by the total amount of attempts ($M = 28.07$, $SD = 3.17$). These efficiency and exploration scores were used for the analyses.

2.2.2. Executive control

To assess executive control, two computer tasks were used. Flanker Fish was used to measure attentional control (Rueda, Posner, & Rothbart, 2005) and Hearts and Flowers was used to measure action control (Davidson, Amso, Anderson, & Diamond, 2006). Each task consisted of three blocks: congruent, incongruent, and mixed. Before each block there was instruction and practice, except for the mixed block of the Hearts and Flowers task. Children were given 4000 ms to respond on the Flanker Fish task and 1500 ms on the Hearts and Flowers task (following Davidson et al., 2006). Responses faster than 200 ms were omitted from the analyses (1.6% for the Flanker Fish task and 1.1% for the Hearts and Flowers task), because these were considered anticipatory (Davidson et al., 2006). Correct responses were assigned a score of one and incorrect responses or no response at all were assigned a score of zero.

2.2.2.1. Attentional control. Flanker Fish measured attentional control, because children had to resolve conflict from competing stimulation (Rueda et al., 2005). Children had to feed the hungry fish that appeared on the screen by pressing a button on the same side the hungry fish were facing (left or right). The hungry fish was usually accompanied by four other fish. There were three possible configurations: the hungry fish swam to the same side as the flankers, the hungry fish swam to the opposite side of the flankers, or the hungry fish swam alone. The first block consisted of blue fish, of which the middle one was hungry. The children had to focus on the middle fish, while inhibiting the tendency to respond to the flankers. The second block consisted of pink fish, of which the flanker fish were hungry. Blue and pink fish alternated in the third block. The children had to switch between the rules that applied to the color of the fish. The first item of a block was not analyzed. This resulted in 16 items for the first two blocks and 44 items in the third block. The Flanker Fish task was reliable, Cronbach's $\alpha = 0.91$, Guttman's $\lambda_2 = 0.92$.

2.2.2.2. Action control. Action control was measured in the second task: Hearts and Flowers. A heart or a flower appeared either on the left or the right side of the screen. In the first block, hearts were presented. Children had to press the button on the same side as the heart appeared. The children's performance showed ceiling effects ($M = 11.01$, $SD = 1.40$). This block was omitted from all further analyses. Flowers were presented in the second block. Children had to press the button on the opposite side as the flower appeared. In the mixed block, hearts and flowers alternated. The first item of the mixed block was not analyzed. This resulted in 12 items for the first and second block, and 32 items in the third block, mixed. The Hearts and Flowers task was reliable, Cronbach's $\alpha = 0.85$, Guttman's $\lambda_2 = 0.86$.

2.2.2.3. Aggregated executive control measures. To confirm that the two tasks measured different aspects of control, a principal components analysis (PCA) was conducted. Oblimin rotation was used to improve the interpretability of the components. The total number of correct responses per block were entered in the PCA, i.e., Flanker Fish block 1 ($M = 11.67$, $SD = 4.02$), block 2 ($M = 10.86$, $SD = 3.51$), and block 3 ($M = 29.40$, $SD = 8.36$), and Hearts and Flowers block 2 ($M = 8.92$, $SD = 2.91$) and block 3 ($M = 21.07$, $SD = 6.35$). The PCA revealed two components; one for each task. The first component (attentional control) showed high loadings in the pattern matrix on the blocks of the Flanker Fish task (0.71–0.84) and it explained 51.58% of the variance. The second component (action control) showed high loadings in the pattern matrix on the blocks of Hearts and Flowers task (0.89–0.90) and it explained 19.32% of the variance. These components scores were used for the analyses.

2.2.3. Verbal and nonverbal skills

2.2.3.1. Nonverbal reasoning. The exclusion task (Bleichrodt, Drenth, Zaal, & Resing, 1987) was used to assess nonverbal reasoning. The stimuli consisted of four abstract figures, of which three belonged to one category. The children had to point to the figure that did not belong to the category. The category could be determined by inducing the rule that underlies the similarity between the figures. To identify the non-member, deductive reasoning was needed based on the underlying rule. When four items were responded to incorrectly, testing was discontinued. There were 30 items, which increased in difficulty. The score was the sum of correct responses. The exclusion task was reliable with Cronbach's $\alpha = 0.82$, and Guttman's $\lambda_2 = 0.84$.

2.2.3.2. Vocabulary. Vocabulary was assessed using the verbal meaning task (Bleichrodt et al., 1987). The child had to choose one out of four figures, which resembled the word spoken by the experimenter. When four items were responded to incorrectly, testing was discontinued. There were 40 items, which increased in difficulty. The score was the sum of correct responses. This task was reliable with Cronbach's $\alpha = 0.84$, and Guttman's $\lambda_2 = 0.85$.

2.3. Procedure

The children were all tested individually in a quiet place within the schools. The tasks were administered in three sessions, each lasting 10–20 min. One session consisted of the Hippo app with game-plays being presented in a random order. In another session, executive control was measured. Nonverbal reasoning and vocabulary were assessed in another session. The sessions were administered in a random order by three experimenters with each child being tested by the same experimenter.

3. Results

3.1. Exploration and efficiency scores on the game-plays of the hippo app

Table 1 summarizes the mean scores of the game exploration and game efficiency score, as well as the descriptive statistics of the cognitive factors. Note that scores on attentional and action control could be negative, because components scores were used. The first research question focused on the in-game behaviors, game exploration and game efficiency, as a function of game-play and level.

To examine the efficiency scores on the three game-plays of the Hippo app a repeated-measures ANOVA on the game efficiency score was conducted with game-play (Slides, Seesaw, and Pendulum) and level (1, 2, 3, and 4) as within-subject factors. Whenever sphericity could not be assumed, according to Mauchly's *W*, the degrees of freedom were corrected using the Huynh-Feldt method. We found an interaction between game-play and level, $F\,(5.70, 421.40) = 13.47$, $p < 0.001$, $\eta^2_p = 0.15$ ($W = 0.62$, $p = 0.021$). This interaction was further investigated by comparing the game-plays per level. Post-hoc analysis with Bonferroni-correction showed that performance on the Slides was higher than on the Pendulum, except for level 4, $p = 0.798$. This effect can be explained by the final level of the slides. A single variable was investigated, namely the surface of the Slides. This variable turned out to be most challenging conceptually. Therefore, a second ANOVA was conducted without level 4 to investigate the effects of game-play and level on the game efficiency score. Now the interaction between game-play and level was not significant, $F\,(3.59, 265.77) = 2.42$, $p = 0.056$, $\eta^2_p = 0.03$ ($W = 0.65$, $p < 0.001$). This effect was near significance, due to the increased performance on level 3 of the Pendulum, see Fig. 2. There were main effects of game-play, $F\,(1.85, 136.79) = 28.97$, $p < 0.001$, $\eta^2_p = 0.28$ ($W = 0.89$, $p = 0.016$), and level, $F\,(1.85, 136.58) = 45.45$, $p < 0.001$, $\eta^2_p = 0.38$ ($W = 0.89$, $p = 0.015$), see Fig. 2. Both effect sizes were large (Cohen, 1988).

Post-hoc analysis with Bonferroni-correction showed that the game efficiency score was higher on the Slides than both the Seesaw, $p = 0.001$, and the Pendulum, $p < 0.001$. The game efficiency score was also higher for the Seesaw than the Pendulum, $p = 0.003$. The main effect of level was linear, $F\,(1, 74) = 88.28$, $p < 0.001$, indicating that game efficiency increased with level. Post-hoc analysis with Bonferroni-correction showed that performance was higher on level 3 than on level 2, $p < 0.001$, and level 1, $p < 0.001$. The performance on level 2 was higher than on level 1, $p = 0.003$. The repeated-measures ANOVA showed that both game-play and level had an effect on game efficiency. These results were expected based on the design of the Hippo app.

Next, a repeated-measures ANOVA was also conducted on the exploration score with game-play (Slides, Seesaw, and Pendulum) and level (1, 2, and 3) as within-subject factors. There was no interaction between game-play and level, *F* (6,

Table 1
The descriptive statistics of the in-game behaviors, exploration and efficiency, and the components of executive control, attentional control and action control, and nonverbal reasoning and vocabulary.

	M	*SD*	Minimum	Maximum
Game exploration score	0.09	0.02	0.03	0.15
Game efficiency score	0.24	0.10	0.06	0.48
Attentional control	0	1	−2.08	1.78
Action control	0	1	−2.53	1.59
Nonverbal reasoning	19.03	5.11	5	29
Vocabulary	27.59	5.99	12	37

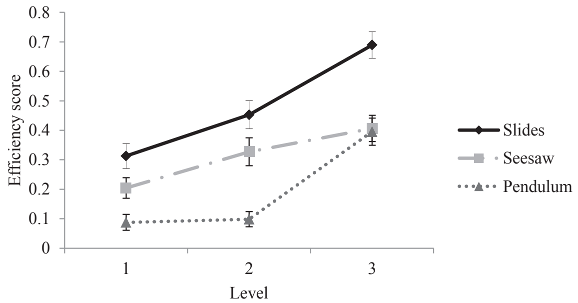

Fig. 2. The mean game efficiency score per game-play (Slides, Seesaw, and Pendulum) over the first three levels per game-play. The error bars represent the standard error of the mean.

296) = 0.53, $p = 0.712$, $\eta^2_p = 0.01$, and there was no effect of game-play, $F(2, 148) = 0.68$, $p = 0.508$, $\eta^2_p = 0.01$. Level did affect the exploration score, $F(6, 148) = 14.70$, $p < 0.001$, $\eta^2_p = 0.17$. Post-hoc analysis with Bonferroni-correction revealed that the exploration score was higher for level 1 compared to level 2, $p = 0.001$, and level 1 compared to level 3, $p = 0.001$. The exploration score of level 2 and level 3 did not differ significantly, $p = 0.258$.

3.2. Individual variation in game exploration and game efficiency

The second research question focused on the relation between exploration and efficiency, the two scores of gaming behavior, and other cognitive factors. In Table 2, it can be seen that the correlation between exploration and efficiency was 0.09. We also tested other curves, but they also showed that game exploration and game efficiency were not significantly related, such as the quadratic curve, $F(2, 72) = 1.38$, $p = 0.259$, and the cubic curve, $F(3, 71) = 0.91$, $p = 0.441$. Therefore we studied the individual differences separately for game exploration and game efficiency. The cognitive factors correlated with both gaming scores, except for action control, see Table 2.

Next, mediation analyses were conducted on the game exploration and efficiency scores. Attentional and action control, as part of executive control, were hypothesized to be the basis on which other skills are acquired. They were, therefore, the independent variables, and nonverbal reasoning and vocabulary were the mediators. The mediation effect was considered significant when the 95% bias-corrected confidence interval (*CI*) did not include zero. This effect was calculated using 50000 bootstrap-samples. The other effects were calculated using ordinary least squares regression (Hayes, 2013). Unstandardized coefficients were used.

In both mediation models, there was a significant indirect effect. Regarding game exploration, vocabulary mediated the effects of attentional control and action control, ab = 0.0039, *CI* = [0.0012, 0.0085], see Fig. 3. In contrast, regarding game efficiency, nonverbal reasoning mediated the effects of attentional control and action control, ab = 0.0078, *CI* = [0.0003, 0.0233], see Fig. 4. Attentional control related significantly to both vocabulary, a = 2.3068, $t(68) = 3.24$, $p = 0.002$, and nonverbal reasoning, a = 1.6571, $t(68) = 2.68$, $p = 0.010$, while action control did not, $p = 0.171$ and $p = 0.225$, respectively. Vocabulary did relate to game exploration, b = 0.0017, $t(66) = 3.23$, $p = 0.010$, while nonverbal reasoning did not, $p = 0.454$. Nonverbal reasoning's relation with game efficiency was near significant, b = 0.0047, $t(66) = 1.80$, $p = 0.076$. While this path of nonverbal reasoning to game exploration was not significant, the mediation effect via nonverbal reasoning was significant, because it also included the paths from attentional and action control to nonverbal reasoning. Vocabulary did not relate to game efficiency, $p = 0.263$.

4. Discussion

The present study set out to investigate game exploration and game efficiency in relation to executive control, vocabulary, and nonverbal reasoning. Our first result was that game exploration depended on the level of the game-play only and that game efficiency related to both the type of game-play and the level of the game-play, as expected. It is therefore feasible to assess children's exploration and efficiency in scientific thinking using a serious game. In addition, we found the scores for game exploration and efficiency were not significantly related. Therefore, we investigated individual differences separately. Our final result was that attentional control predicted both game exploration and game efficiency. We found the relation between game exploration and attention control to be mediated by vocabulary, and the relation between game efficiency and attention control by nonverbal reasoning.

The first result was that the in-game behaviors were related to the level of the game-play. There was less exploration and more efficiency when the level increased. An increase in level meant a decrease in the number of variables. When there are fewer variables, it seemed easier to solve the game-play. This result is in line with other studies on scientific reasoning that found that performance depends on the number of variables. When kindergartners have to predict and explain what side of a balance beam would go down, their performance decreases when they have to incorporate distance (to the fulcrum), besides

Table 2
Pearson's *r* correlations between game exploration, game efficiency, executive control (attentional and action control), nonverbal reasoning, and vocabulary.

	1	2	3	4	5	6
1 Game exploration score	1					
2 Game efficiency score	0.09	1				
3 Attentional control	−0.03	0.32**	1			
4 Action control	0.15	0.07	0.44**	1		
5 Nonverbal reasoning	0.14	0.39**	0.44**	0.23*	1	
6 Vocabulary	0.37**	0.35**	0.50**	0.24*	0.55**	1

Note. $^{**}p < 0.01$, $^{*}p < 0.05$.

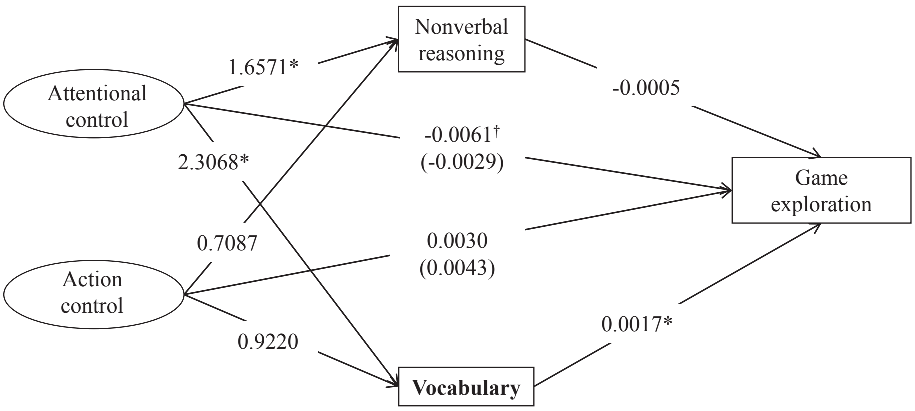

Fig. 3. Mediation model of executive control, nonverbal reasoning, vocabulary, and game exploration. Total effects (i.e., c′) are between brackets. The significant mediator is printed in bold. The mediation model was significant, $R^2 = 0.17$, $F\,(4, 66) = 3.34$, $p = 0.015$. *Note*. $^{*}p < 0.05$, † $p < 0.10$.

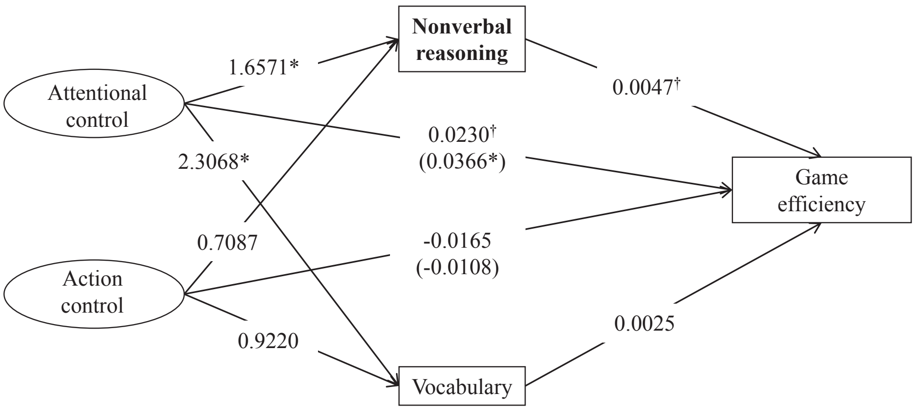

Fig. 4. Mediation model of executive control, nonverbal reasoning, vocabulary, and game efficiency. Total effects (i.e., c′) are between brackets. The significant mediator is printed in bold. The mediation model was significant, $R^2 = 0.20$, $F\,(4, 66) = 4.07$, $p = 0.005$. *Note*. $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

weight, in their predictions and explanations (Siegler, 1976). Another study on scientific reasoning of kindergartners found that children design fewer experiments correctly when the number of variables they have to set increases (Van der Graaf et al., 2015). The level of the game-play also related to game exploration. When there were fewer variables, there was less exploration. This is in line with the design of the game-plays. When there are fewer variables, there are fewer possibilities to interact with the game, which results in less exploration. Also, the levels with fewer variables were easier to solve, as revealed by the effect of game efficiency. Another aspect of the app was that there were three different game-plays, the slides, the seesaw, and the pendulum. The present results showed that the game-play of the slides was easiest, followed by the seesaw, and that the pendulum was most difficult. The slides were expected to be easiest, because conceptual understanding of the slides is present in kindergartners (Inhelder & Piaget, 1958). Inhelder and Piaget (1958) also described the development of understanding on the seesaw, which, just as Siegler (1976) showed, is difficult for kindergartners, as they experience difficulties in identifying distance to the fulcrum as a variable. In their descriptions of the physics topics, Inhelder and Piaget (1958) made clear that complete understanding of the pendulum emerges latest.

Furthermore it is interesting to note that exploration and efficiency turn out to represent independent game behaviors. Although it has been proposed that exploration resolves ambiguity, seemingly in order to start discovering explanations for unusual or unexpected events (Legare, 2014), the present results showed that there can be exploration without the need for solving the problem (or explaining what is happening). In a similar vein, there can be efficient behavior without attempting to explore all possibilities.

With respect to individual variation, game exploration was related to attentional control and vocabulary. Attentional control can be relevant for exploration, since it enhances formulation of strategies necessary to try out differential aspects of the game (Fernyhough & Fradley, 2005) and to selectively process stimuli (Hopfinger, Buonocore, & Mangun, 2000). The effect of attentional control on game exploration was indirect through vocabulary, which can be explained by the fact that vocabulary may stimulate talking about science (Snow, 2010). In the current study, it might have aided in the formulating of an exploration strategy or in the representation of new concepts during exploration (Werker et al., 2002).

With regard to game efficiency, both attentional control and nonverbal reasoning were involved. The fact that game efficiency requires the coordination and combination of several aspects of the game, such as settings of the variables and previous outcomes explains that the game was likely to be cognitively challenging and thus demanding of resources of the attentional control system to maintain goal-directed behavior (Eysenck, Derakshan, Santos, & Calvo, 2007). Attentional control can also be seen as highly relevant for the generation of the right problem-solving strategies during game behavior (Fernyhough & Fradley, 2005). Besides attentional control, nonverbal reasoning related to game efficiency. Many types of reasoning have been found to be involved in scientific thinking, such as nonverbal reasoning (Van der Graaf et al., 2015), syllogistic reasoning (Wagensveld et al., 2015), and general fluid ability (Mayer, Sodian, Koerber, & Schwippert, 2014). As Dunbar and Klahr (2012) have shown, reasoning helps in extrapolating rules to yet-to-be observed phenomena in order to generate a response, and that the inference of rules requires reasoning in order to adjust response to the problem to be solved.

What is left for discussion is the absence of significant relations between action control and the in-game behaviors, between nonverbal reasoning and game exploration, and between vocabulary and game efficiency. Regarding action control, it has been proposed that it could help in stopping automatic responses (Ponitz et al., 2008). However, it seems that the discovery of physics was an engaging task, which does not stimulate automatically responding. This notion is in line with the finding of Van de Sande, Segers, and Verhoeven (2015) that autonomous and explorative games can engage children for relatively long time periods, before they tend to go off task. They found off-task behavior to be related to action control. For nonverbal reasoning, it makes sense that this is not directly needed to explore a game, since exploration is unconstrained and would therefore need no rules or formulation of rules. In contrast, exploration does require mental representations of new scientific concepts, in which vocabulary can aid (Snow, 2010). Game efficiency was about correctly responding in few attempts. In order to solve the physics problems, there was no need to verbalize the variables and/or their effects. The children had to use their own logic to set the variables and to adjust responses given the observed effects (Dunbar & Klahr, 2012). This is not to say that language does not play a role in scientific thinking, as vocabulary was correlated with game-efficiency. An interesting finding is that vocabulary and nonverbal reasoning did correlate with action control, which can be explained by the supportive role of action control in task performance. Action control allows children to control their action, so that children can behave according to the task demands. However, these relations were no longer significant in the mediation models, where attentional control was controlled for. Attentional control, therefore, seems to be a more relevant factor in vocabulary and nonverbal reasoning, which can be explained by the fact that most children can control their actions.

A limitation on the present study is that the initial and subsequent knowledge of the physics topics have not been assessed. While kindergartners' in-game behavior was decomposed, their learning gains were not investigated. Learning gains might be present if the game is used as an intervention tool to teach children about the process of discovery and manipulating variables, and about the underlying physics by observing the simulated result. It would be interesting to study how much children learn from this game and how experience with the game might affect learning gains. Another limitation is that previous experience with gaming and/or with tablet use was not assessed. The children were introduced to the tablet and to the game, but this might have been easier to understand for children with higher levels of digital literacy. Next to including these measures, a suggestion for future research would be to include a measure of curiosity, as curiosity can drive exploration, even when apparently more urgent situations would require a (efficient) response (Berlyne, 1966).

An important implication for educational practice is that children, as young as four to six years old, can use a serious game to discover the laws of physics. This way, kindergartners can already be engaged in science education. A serious game can provide a rich and informal learning environment for these young children (cf. Morris et al., 2013). Such an environment may challenge the children to manipulate variables and observe the results in order to explore the effects of the variables and/or to solve the problems. This confirms that exploration and efficiency can be assessed with a serious game (cf. Legare, 2014). Another practical implication is that it can be assessed how children at kindergarten level approach a serious game about physics. The present study made it clear that children differed in the way they discovered the laws of physics and also in their efficiency in doing so. This supports the notion that both exploration and efficiency can be assessed in an informal scientific discovery learning context (cf. Bonawitz et al., 2012).

Two implications followed from the design of the game. Three different game-plays were used, namely the slide, the seesaw, and the pendulum. Each had three levels, which differed in the number of variables. Since no interaction was found between game-play and level, it means that the same cognitive principles hold between the game-plays. Whenever the number of variables changed (i.e. a change in level), it affected performance, no matter what type the game-play was. The second implication is that variables should be introduced gradually. When all variables were free to manipulate, performance was relatively low. This suggests that children should be introduced to the variables one by one, so they can be encoded and used in their discovery.

To summarize, game exploration related to the level of game-play, and game efficiency related to both the level of game-play as the game-play (slides, seesaw, and pendulum). From a scientific thinking point of view, this study was the first to relate exploration and efficiency to each other and to other cognitive factors. From a developmental point of view, this study extended the increasing acknowledgement of young children's abilities. To conclude, serious games can be used to investigate how children approach scientific activities, i.e., exploration versus efficiency. Attentional control is a key cognitive factor in exploration via vocabulary and to efficiency via nonverbal reasoning in discovering the laws of physics with a serious game.

Acknowledgements

This research was funded by a grant from the Dutch Curious Minds Research Program.

References

Berlyne, D. M. (1966). Curiosity and exploration. *Science, 153*, 25–33. http://dx.doi.org/10.1126/science.153.3731.25.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1987). *Revision Amsterdam child intelligence test. Manual*. Lisse, The Netherlands: Swets & Zeitlinger.

Bonawitz, E. B., Van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*, 215–234. http://dx.doi.org/10.1016/j.cogpsych.2011.12.002.

Cartwright, K. B. (2012). Insights from cognitive neuroscience: the importance of executive function for early reading development and education. *Early Education and Development, 23*, 24–36. http://dx.doi.org/10.1080/10409289.2011.615025.

Cecil, L. M., Gray, M. M., Thornburg, K. R., & Ispa, J. (1985). Curiosity-exploration-play-creativity: The early childhood mosaic. *Early Child Development and Care, 19*, 199–217. http://dx.doi.org/10.1080/0300443850190305.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078. http://dx.doi.org/10.1016/j.neuropsychologia.2006.02.006.

Dede, C. (2010). Comparing frameworks for 21st century skills. In J. A. Bellanca, & R. Brandt (Eds.), *21st Century skills: Rethinking how students learn*. Bloomington, IN: Solution Tree Press.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. http://dx.doi.org/10.1146/annurev-psych-113011-143750.

Dunbar, K. N., & Klahr, D. (2012). Scientific thinking and reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning. Oxford Handbooks Online*. http://dx.doi.org/10.1093/oxfordhb/9780199734689.013.0035.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*, 336–353. http://dx.doi.org/10.1037/1528-3542.7.2.336.

Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: Relations with task difficulty and task performance. *Cognitive Development, 20*, 103–120. http://dx.doi.org/10.1016/j.cogdev.2004.11.002.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 5*, 28–45. http://dx.doi.org/10.14786/flr.v2i3.96.

Greenfield, P. M., Camaioni, L., Ercolani, P., Weiss, L., Lauber, B. A., & Perucchini, P. (1994). Cognitive socialization by computer games in two cultures: Inductive discovery or mastery of iconic code? *Journal of Applied Developmental Psychology, 15*, 59–85. http://dx.doi.org/10.1016/0193-3973(94)90006-X.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii international conference on system sciences, Waikoloa, HI. January 6-9, 2014*. http://dx.doi.org/10.1109/HICSS, 2014.377.

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology, 98*, 307–326. http://dx.doi.org/10.1037/0022-0663.98.2.307.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*, 284–291. http://dx.doi.org/10.1038/72999.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures (A. Parsons & S. Milgram, Trans.)*. New York, NY: Basic Books.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48. http://dx.doi.org/10.1207/s15516709cog1201_1.

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science, 333*, 971–975. http://dx.doi.org/10.1126/science.1204528.

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development, 23*, 472–487. http://dx.doi.org/10.1016/j.cogdev.2008.09.007.

Kuhn, D. (2004). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The Blackwell handbook of childhood cognitive development. Blackwell Reference Online*. http://dx.doi.org/10.1111/b.9780631218418.2004.00020.x.

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In W. Damon, & R. M. Lerner (Eds.), *Child and adolescent development: An advanced course*. Hoboken, NJ: John Wiley & Sons. http://dx.doi.org/10.1002/9780470147658.chpsy0222.

Kwon, Y., & Lawson, A. E. (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching, 37*, 44–62. http://dx.doi.org/10.1002/(SICI)1098-2736(200001)37:1<44::AID-TEA4>3.0.CO;2-J.

Lefevre, J., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S., et al. (2013). The role of executive attention in the acquisition of mathematical skills for children in grades 2 through 4. *Journal of Experimental Child Psychology, 114*, 243–261. http://dx.doi.org/10.1016/j.jecp.2012.10.005.

Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives, 8*, 101–106. http://dx.doi.org/10.1111/cdep.12070.

Lorch, R. F., Lorch, E. P., Freer, B. D., Dunlap, E. E., & Hodell, E. C. (2014). Using valid and invalid experimental designs to teach the control of variables strategy in higher and lower achieving classrooms. *Journal of Educational Psychology, 106*, 18–35. http://dx.doi.org/10.1037/a0034375.

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relation with cognitive abilities. *Learning and Instruction, 29*, 43–55. http://dx.doi.org/10.1016/j.learninstruc.2013.07.005.

Morris, B. J., Croker, S., Zimmerman, C., Gill, D., & Romig, C. (2013). Gaming science: The "gamification" of scientific thinking. *Frontiers in Psychology, 4*, 1–16. http://dx.doi.org/10.3389/fpsyg.2013.00607.

Osborne, J. (2013). The 21st century skills challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity, 10*, 265–279. http://dx.doi.org/10.1016/j.tsc.2013.07.006.

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology, 31*, 153–179. http://dx.doi.org/10.1111/j.2044-835X.2012.02082.x.

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., McDonald Conner, C., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*, 141–158. http://dx.doi.org/10.1016/j.ecresq.2007.01.004.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28*, 573–594. http://dx.doi.org/10.1207/s15326942dn2802_2.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520. http://dx.doi.org/10.1016/0010-0285(76)90016-5.

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development, 71*, 26–35. http://dx.doi.org/10.1111/1467-8624.00115.

Snow, C. E. (2010). Academic language and the challenge of reading. *Science, 328*, 450–452. http://dx.doi.org/10.1126/science.1182597.

Van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: Dynamic assessment of the control of variables strategy. *Instructional Science, 43*, 381–400. http://dx.doi.org/10.1007/s11251-015-9344-y.

Van de Sande, E., Segers, E., & Verhoeven, L. (2015). The role of executive control in young children's serious gaming behavior. *Computers & Education, 82*, 432–441. http://dx.doi.org/10.1016/j.compedu.2014.12.004.

Vosniadou, S. (2002). On the nature of naïve physics. In M. Limón, & L. Mason (Eds.), *Reconsidering conceptual change* (pp. 61–76). http://dx.doi.org/10.1007/0-306-47637-1_3.

Wagensveld, B., Segers, E., Kleemans, T., & Verhoeven, L. (2015). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, 1–15. http://dx.doi.org/10.1007/s11251-014-9334-5.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1–30. http://dx.doi.org/10.1207/S15327078IN0301_1.