Feasibility, enjoyment, and language comprehension impact of a tablet- and GameFlow-based story listening game for kindergarteners: a methodological and mixed method study

Femke Vanden Bempt[1,2], Maria Economou[2,1], Ward Dehairs[2], Maaike Vandermosten[2], Jan Wouters[2], Pol Ghesquière[1], Jolijn Vanderauwera[3,4,1,2]

Affiliations:

[1] KU Leuven, Faculty of Psychology and Educational Sciences, Parenting and Special Education Research Unit, B-3000 Leuven, Belgium

[2] KU Leuven, Department of Neurosciences, Research Group ExpORL, B-3000 Leuven, Belgium

[3] Psychological Sciences Research Institute, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

[4] Institute of Neuroscience, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

Abstract

Background: Enjoyment plays a key role in the success and feasibility of serious gaming interventions. Unenjoyable games will not be played and in the case of serious gaming, learning will not occur. Given this importance, a so-called GameFlow model has been developed, intending to guide (serious) game developers in the process of creating and evaluating enjoyment in digital (serious) games. Regarding language learning, a variety of serious games, targeting specific language components, exist on the market, albeit often without available assessments of enjoyment or feasibility.

Objective: The current study evaluates enjoyment and feasibility of a tablet-based serious story listening game for kindergarteners, developed based on the principles of the GameFlow model. Additionally, this study preliminarily explores the possibility of the game to foster language comprehension.

Methods: Within the framework of a broader preventive reading intervention, 91 five-year old kindergarteners at cognitive risk for dyslexia were asked to play the story game for 12 weeks, six days per week, either combined with a tablet-based phonics intervention or active control games. The story game mainly involved story listening and rating, and responding to content-related questions. Game enjoyment was assessed via post-intervention questionnaires, a GameFlow-based evaluation, and in-game story rating data. Feasibility was determined based on in-game general question response accuracy (QRA), reflecting the difficulty level, attrition rate, and final game exposure and training duration. Additionally, in order to investigate whether game enjoyment and difficulty could influence feasibility, final game exposure and training duration were predicted based on in-game initial story ratings and initial QRA. The possible growth in language comprehension was explored by analyzing in-game QRA as a function of game phase and baseline language skills.

Results: Eventually, data from 82 participants were analyzed. Questionnaire and in-game data suggested an overall enjoyable game experience, but the GameFlow-based evaluation implied room for improving the game design. General QRA confirmed a well-adapted difficulty level for the target sample. Moreover, despite an overall attrition rate of 32 participants (39%), 74 participants (90%) still completed 80% of the game, albeit with a large variation in training days. Higher initial QRA resulted in a significantly higher game exposure (β=0.35; P<.001) and lower initial story ratings significantly slackened the training duration (β=-0.16; P=.003). In-game QRA was positively predicted by game phase (β=1.44; P=.004), baseline listening comprehension (β=1.56; P=.002), and vocabulary (β=0.16; P=.01), with larger QRA growths over game phases in children with lower baseline listening comprehension skills (β=-0.08; P=.04).

Conclusions: Generally, participants experienced the story game as enjoyable and feasible. Yet, the GameFlow-model evaluation and predictive relationships imply room for further game design improvement. Furthermore, our results cautiously suggest a potential of the game to foster language comprehension, but future randomized controlled trials need to elucidate the actual gaming impact on language comprehension.

Keywords: Serious gaming; language comprehension; enjoyment; feasibility; GameFlow

Introduction

Background

Over the past two decades, there has been a major increase in the development of digital serious games for children with special educational needs [1,2]. This clear trend presumably relates to the popularity and motivational aspects of non-educational digital games and the rising exposure to technological devices in children's current daily lives [3]. Within the field of education, digital serious games are described as games for electronic devices with certain game rules and player outcomes, which aim at entertaining, but pursue at least one additional learning objective (e.g., fostering mathematics, language, reading, etc.) [2,4]. In order to obtain the desired educational impact with such games, children are often required to engage in the intervention on a recurrent basis and for a considerable amount of time. Hence, game enjoyment, which triggers gameplay and increases the feasibility to maintain playing [5,6], is a key to the success of serious gaming interventions [7]. When evaluating serious games, it is thus of high relevance to assess not only the educational efficacy of interest, but also these related aspects of game enjoyment and feasibility [2]. The current study will address enjoyment and feasibility of a newly designed serious story listening game suitable for five-year old children. Additionally, given the focus of the game on story listening, the current study aims at preliminarily investigating its potential to train language comprehension [8].

Language comprehension is the process of interpreting spoken language input [9]. It involves speech processing at the basic auditory level, derivation of word meaning, knowledge of syntactic and morphological structures, and integration of information within a broader language context [10]. Language comprehension creates opportunities for daily-life verbal communication and lays the foundations for reading comprehension long before

receiving reading instruction [11,12]. Hence, unsurprisingly, language comprehension forms a well-known precursor for academic achievement [8,13,14], employment [14,15], and social participation [14,16]. Individual differences in language comprehension already exist at an early age and remain relatively stable over time [17]. Hence, dramatically, without remediation, young children with poor comprehension are more likely to lag behind their peers throughout their entire development [17].

Problems with language comprehension are apparent in a high proportion of children with a developmental language disorder (DLD) [18]. DLD is characterized by severe problems with expressive and/or receptive language development in the absence of a clear-cut neural, cognitive, or auditory cause, impacting daily life communication [19]. Remarkably, children diagnosed with DLD who experience comprehension deficits more often have persistent language problems [18], respond less well to intervention, and require more extensive support than children with DLD who exhibit language deficits in the expressive domain only [20]. Next to DLD, lower comprehension skills have also been observed in second language learners [21], in children with low socio-economic status (SES) environments [22], in hearing impaired children [23], and in children diagnosed with developmental dyslexia [24,25].

Language comprehension deficiencies benefit from interventions early in childhood [14]. Indeed, a recent systematic scoping review of Tarvainen and colleagues [18] revealed that comprehension in young children is effectively tackled by (1) guidance in the parent-child or clinician-child communication strategies, or (2) by targeting receptive and expressive language aspects, such as vocabulary, morphology, and inferential language. Moreover, in young children under the age of six, implicit therapy techniques, in which the child is exposed to optimal language without explicitly explaining certain rules, are preferred over explicit

instructional approaches [26]. An example of a rather implicit approach concerns the method of storybook listening.

Given the evidence of storybook listening to foster several language comprehension components in young children, e.g., receptive vocabulary, morphosyntax, and narrative comprehension [27–29], both in physical [30–33] or digital settings [34–35], to the best of our knowledge, no study has ever embedded the story listening method in a serious gaming context. Often, digital games offer stimulating audio-visual game worlds and appealing rewards [36,37]. Given these motivational aspects [38,39], serious gaming is already widely applied in educational, psychological, and medical contexts [40–44], including the field of language learning (e.g., game-based interventions specifically targeting vocabulary [45]). However, when designing a commercial and/or serious digital game, a detailed knowledge of game enjoyment is of crucial relevance. A digital game that is not experienced as fun, will not be played [5] and consequently, in the case of serious gaming, learning will not occur. Hence, when evaluating serious games, both aspects of enjoyment as well as educational impact must be considered [2]. Yet, despite its high relevance, the factor of enjoyment is hardly ever evaluated in existing serious games due to the lack of proper frameworks. To both clarify the concept of game enjoyment as well as to facilitate its evaluation, Sweetser and Wyeth [5] proposed a summarizing framework, i.e., the GameFlow model, which is based on the extensive gaming literature and the general theory of Flow [46]. The model intends to guide serious game developers in the process of creating and improving games that are both educational and enjoyable [6]. More specifically, the GameFlow model [5] proposes eight interrelated gaming elements, which are important to attain overall game enjoyment: (1) concentration, (2) challenge, (3) player skills, (4) control, (5) clear end or intermediate goals, (6) feedback, (7) immersion, and (8) social interaction. The concentration principle states that a game becomes enjoyable when a player is able to concentrate on the game itself without

being constantly distracted by surrounding background factors such as game and volume settings, etc. The challenge principle argues that, in order to enjoy a game, it must be sufficiently challenging. The player's skills principle reasons that a game should provide opportunities to improve playing skills at a pace that is adjusted to the capacities of the individual player. The control principle asserts that players must experience a feeling of control over their actions, decisions, the appearance of their avatars and strategies, and the game world. The fifth clear end or intermediate goal principle states that a game must inform the player of clear end or intermediate goals from the start. The principle of feedback emphasizes the importance of appropriate feedback about game progress or errors at appropriate times. The immersion principle states that players should be completely absorbed by the game-world. Finally, the social interaction principle underlines the relevance of social interaction during gameplay. To successfully embed each GameFlow element in a digital game, the GameFlow model [5] proposes a set of predefined implementation criteria for each element. For example, one of the criteria to attain the concentration principle states that game developers should not add irrelevant game tasks on top of the main learning task (consider the study of Sweetser and Wyeth [5] for a full overview of the elements and their corresponding criteria). When evaluating games based on these GameFlow criteria, Sweetser and Wyeth [5] were able to successfully distinguish between low- and high-rated games. Thus, the validity of the GameFlow model to evaluate enjoyment in digital games is warranted.

Objectives of the current study

The current research article will address enjoyment and feasibility of a newly developed serious tablet- and game-based story listening intervention (henceforth story game), for which the GameFlow model had served as a guideline in the design process. Moreover, given the focus of the game on story listening, which is a frequently used method to increase young children's language comprehension [8], its potential to foster language

comprehension will also be preliminarily investigated. The story game was originally developed to enhance basic auditory speech processing by modifying the recorded speech signals of the stories with a so-called envelope enhancement (EE) algorithm [47–51] and as such, boost phonology and reading in kindergarteners at cognitive risk for dyslexia. The EE algorithm automatically detects and amplifies important rhythmic acoustic cues of the speech envelope, i.e., onset rise times, which are considered to be important for phonological skill (and potentially reading) development [52,53]. The story game was implemented in a broader preventive reading intervention aiming at (1) evaluating the efficacy of a tablet-based phonics intervention specifically targeting reading and phonology, i.e., GraphoGame Flemish (GG-FL) [54,55], and (2) investigating a possible educational boosting effect of the EE story game on top of GG-FL, which will be addressed in a future study. Story game enjoyment will be evaluated in three ways based on (1) post-intervention child and parental questionnaires, (2) fulfilment of the GameFlow criteria, and (3) in-game data indicative of game enjoyment. Feasibility will be addressed by analyzing the difficulty level of the game, the attrition rate, and the individuals' final game exposure and training duration together with their corresponding gaming profiles. Language comprehension growth will be addressed by exploring the accuracy on in-game story content-related questions as a function of game progress and baseline language subskills important for overall language comprehension, e.g., baseline vocabulary, morphological awareness, and listening comprehension [27–29]. That way, we could test the hypothesis whether language comprehension improved when progressing through the game, when controlling for baseline language skills, but also whether children with lower baseline language skills showed larger growing potentials.

Methods

Participants

Following a school-based screening in the third year of kindergarten (n=1225), 149 five-year old children (119 children with and 30 children without an elevated cognitive risk for dyslexia) enrolled in a game-based preventive reading intervention study (Trial registration number: S60962 – assigned by the Clinical Trial Center of UZ Leuven, Belgium). A cognitive risk for dyslexia was assigned when a child scored above the 10th percentile on a non-verbal reasoning test [56] and below percentile 30 on minimally two out of three assessments of robust reading predictors, i.e., letter knowledge [57], phonological awareness [57,58], and rapid automatized naming [59]. Typically developing children scored above percentile 40 on all reading precursors and were matched to the risk-sample based on non-verbal reasoning ability, school environment, and gender. Consider the studies of Van Herck et al., Vanden Bempt et al., and Verwimp et al. [49,55,60] for a more detailed description of the screening tasks, procedure, and participant selection. All selected participants were in their third year of kindergarten, Flemish monolingual Dutch-speaking, born in 2013, and had a schooling period of minimally 20 months. None of the selected children reported an additional behavioral or familial risk for ADHD, language and/or articulatory problems, a severe hearing impairment, or neurological deficits. Furthermore, since Flemish schools only provide reading instruction from first grade onwards, all participants were considered pre-reading. This so-called pre-reading phase was also confirmed by an unanimous floor-effect on a pre-intervention reading test [55]. Signed informed consents were obtained for all the participants and the study was approved by the Medical Ethical Committee of UZ Leuven, KU Leuven.

Study design and procedure

Of the entire intervention study sample (n=149), 91 at-risk children were asked to play the story game. Within the framework of the reading intervention study, these 91 children

were randomly assigned to one of three experimental groups. The first group (GGFL_EE group, n=31) played the story game with envelope enhanced stories and combined it with the phonics-based GG-FL intervention [54]. For technical details of the EE algorithm, consider the study of Van Herck and colleagues [49]. The second group received the same intervention as the first with the only difference that there was no envelope enhancement applied to the stories in the story game (GGFL_NE group, n=31). The third group also played the non-enhanced story game and combined it with tablet-based commercial Lego- and Duplo-applications (henceforth active control (AC) games), which did not train any reading-related skills (AC_NE group, n=29). The intervention lasted 12 weeks / 84 days in total and took place in the second semester of the last and third year of kindergarten. The remaining 28 at-risk and 30 non-at-risk children of the total sample served as the at-risk passive control and typically developing control group of the reading intervention study respectively and did not receive any type of digital gaming intervention. Since the current study mainly focuses on the story game feasibility and enjoyment, these last two control groups, who did not play the story game, will not be discussed any further in the current research article. Prior to the intervention period, all necessary games, including the story game, were installed on 91 tablets (Samsung Galaxy E9.6). All story sound levels of the story game were calibrated to be played at 60 dB-A over ATH-m20x headphones, using speech-weighted noise that was constructed based on an average spectrum of the different storytellers [61]. Then, a variety of cognitive-linguistic tasks, including baseline language skills (listening comprehension, receptive vocabulary, and morphological awareness) were individually assessed in all participants at school in a quiet test room. After this baseline test-phase, the 91 at-risk children who were assigned to one of the three intervention groups received a tablet, its corresponding calibrated headphone, and a manual for parents with instructions to start the tablet intervention at home. With support of a reward calendar with stickers, they were

instructed to combine story game sessions with 15-minute GG-FL/AC sessions for six days per week over a period of 12 weeks. This equaled a total of 72 gaming sessions, both for GG-FL/AC as well as for the story game. After the intervention, parents and children received a questionnaire assessing motivation, enjoyment, and engagement during the intervention period.

Story game intervention

Overall game description

The tablet-based story game application was developed in Unity 3D [62] and was programmed entirely using the C# scripting language. For an overview of the rationale, developmental design-related decisions, the stories, and programming details of the game, consider the technical story game development guide (Multimedia Appendix 1). The story game provided 72 gaming sessions, each of which mainly consisted of listening to one long (e.g., approximately ten minutes) or two short (e.g., twice approximately five minutes) stories, accompanied with story illustrations. Game settings prevented children from playing more than one session per 24 hours. The 72 sessions were categorized in 18 play phases of four grouped sessions containing four (longer) to eight (shorter) stories from the same book series and author. In total, 87 stories from 14 different book series were implemented in the game. Generally, the game contained three main game modes, i.e., (1) the main intervention task environment, which involved the actual story listening, a story rating, and responding to content-related questions (2) the virtual hub world, in which players consulted their game progress on a map, and (3) an avatar customization system, where players could buy accessories for their game avatars or buy new avatars. Most of the artwork in the game, such as 3D models, textures and animations which were used to create the virtual hub world, the avatars, and their accessories, was either custom-made, acquired from the Unity Asset Store [63], or acquired from child-friendly projects of the Dyslexia Research Collaboration

(DYSCO) team at KU Leuven, such as Diesel-X [58]. The resulting art style was characterized by flat primary colors, avatars with exaggerated cartoon-like proportions, and simplified basic models and animations, suitable for the chosen target age group [64,65].

Main intervention task environment

The main intervention task presented recorded story audio, all in Dutch, with accompanying illustrations (see upper left panel of Figure 1), followed by a story appreciation rating (see upper right panel of Figure 1) and audio questions with a simple multiple-choice system (see lower left panel of Figure 1). Questions were formulated by the research group and examined listening comprehension by recalling information that was explicitly mentioned in the story text. Although not directly assessed, the language level of the questions was intended to match the language and cognitive capacities of five-year old children. The question part put a strong emphasis on feedback, awarding one coin per correct response with added flair in the form of animations and sound effects (see lower right panel of Figure 1). Each session contained three content-related questions. Hence, given the total amount of 72 game sessions, players could maximally earn 216 coins throughout the game.

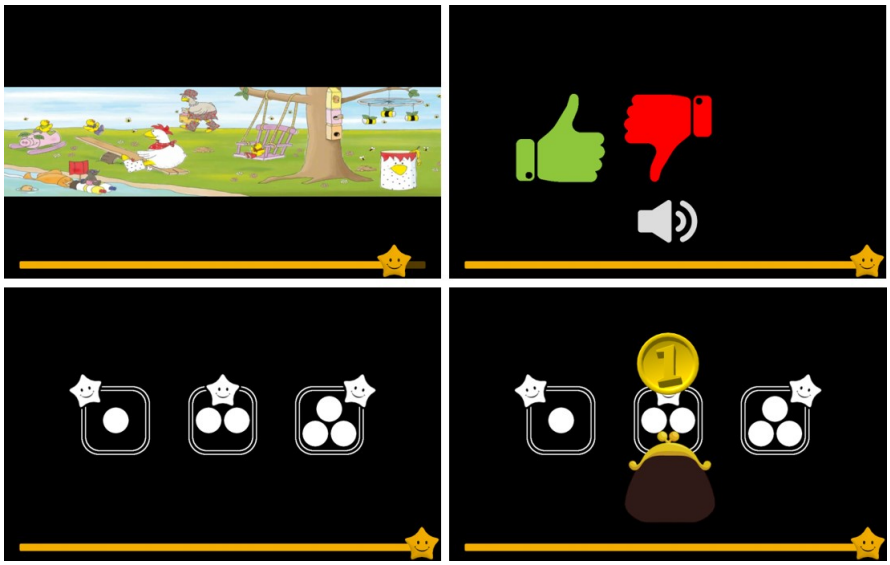


Figure 1. Main intervention task. Upper left panel) Story listening: © 2016 Clavis Publishing, Hasselt-Alkmaar-New York, illustrations from *Lotta ruimt het op* from Diane Put & Rik De

Wulf. Upper right panel) Story rating. Lower left panel) Multiple choice questions. Lower right panel) Reward for each correct response.

Virtual hub world

Similar to classic digital games such as Super Mario Bros, the different game sessions were tied together in a so-called virtual hub world. The virtual hub world provided an overview of each game session in the form of a cylindrical stage. Every time a player successfully completed a game session, his or her avatar jumped from one cylindrical stage to the next, leaving a star on the stage that was just abandoned (see Figure 2). As such, players gained information on their progress throughout the game and could maximally earn 72 stars (i.e., the total amount of game sessions).



Figure 2. Completion of a game session shown in the virtual hub world.

Avatar customization system

To provide added incentives and rewards to keep the player engaged, an avatar customization system was implemented, which featured a virtual store where players could spend coins (Figure 3). These coins were earned through correct responses on the content-related story questions in the main intervention task. The store offered new avatars and accessories for current avatars, which were displayed in the virtual hub world. The items had an emphasis on quantity, to make sure that the player could acquire roughly one new item per story session, without running out long before reaching the end of the story game.

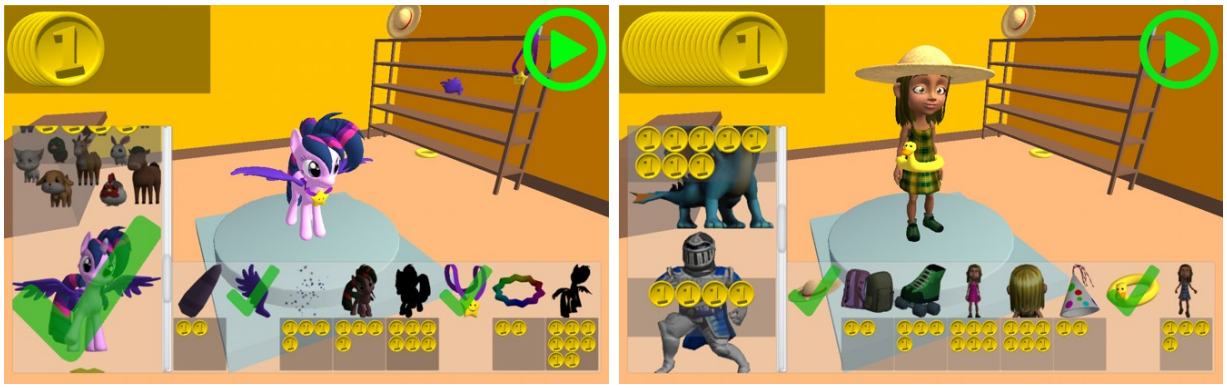


Figure 3. Avatar customization system of the story game.

Story game session

A single game session occurred as follows: first, players watched a short cut-scene of their avatar, positioned on one of the 72 possible stages in the virtual hub world, representing the 72 game sessions and thus, their progress throughout the game. Second, the main intervention task occurred, in which the player was required to actively listen to one or two stories, depending on the particular story session. After listening, the player rated the story based on a green or red thumb (i.e., like or dislike respectively) and received three content-related multiple-choice questions, for which one coin could be earned per correct response. Following the short quiz, the player automatically ended up in the avatar customization system where the earned coins could be spent on accessories for the avatar or to buy extra avatars. Finally, the player was sent back to the hub world in which the new or newly decorated avatar jumped to the subsequent cylindrical stage, while a dancing star appeared on the one that was just abandoned (for a demonstration movie of the story game, see Multimedia Appendix 2). Useful player data of individual player accounts (e.g., game progress, question response accuracy, play dates and hours, and story rating information) were automatically logged on our university server and sent daily to the research group.

Test battery and questionnaires

Baseline listening comprehension

The subtest 'Understanding spoken paragraphs' was adapted from the CELF-4-NL [66] and was used to evaluate baseline listening comprehension. The experimenter read aloud two short stories. Each story was then followed by five content-related questions. The child received one point per correct response. The maximum score of this test was 10.

Baseline receptive vocabulary

Receptive vocabulary was measured using the Peabody Picture Vocabulary Test-III-NL (PPVT-III-NL) [67], containing 17 subtests with 12 test trials each. In each trial, the child heard a target word and was asked to select the corresponding picture out of four alternatives. The test was interrupted when the child made nine or more errors in one subtest. The raw receptive vocabulary score was computed by subtracting the sum of the errors across all administered subtests from the number of the trial that was assessed last. This raw score was then converted into a standard score and represented the final receptive vocabulary score (mean: 100 +/- 15).

Baseline morphological awareness

The 'Word Structure' subtest from the CELF-Preschool 2-NL [68] was used to evaluate morphological awareness. In this task, the ability to correctly apply morphological rules of Dutch, e.g., conjugation, derivation, flexion, pronouns, and degrees of comparisons, was measured. The test contained 23 test items, containing two pictures each. The first picture was fully described by the experimenter (e.g., "This girl eats"/ "Dit meisje eet") followed by an incomplete description of the second picture (e.g., "and this girl …"/ "en dit meisje …"), which had to be completed by applying a certain morphological rule of Dutch (e.g., "and this

girl sleeps"/ "en dit meisje slaapt" – morphological rule: conjugation of a regular verb in third singular form). Based on instructions of the test manual, the test was interrupted when the child made seven consecutive mistakes. The maximum score was 23.

Post-intervention child and parental questionnaires

After the intervention period, parents independently filled out a short questionnaire including three questions related to motivation, encouragement, and sustained attention during story gameplay. Accompanied by a member of the research group at school, intervention enjoyment was also measured in all children using two components of the Fun Toolkit survey instrument [69], which was developed to gather children's opinions on technology. On the one hand, children were asked to assign story game enjoyment on a five-point-Likert-scale-based Smiley-o-meter, in which the five scales were represented by smileys (ranging from a very unhappy smiley meaning 'I did not like the game at all' to a very happy smiley meaning 'I liked the game very much'). On the other hand, the participants were asked whether they were willing to redo the intervention (see Multimedia Appendix 3 for an overview of the five child and parental categorical questions and their response possibilities).

Statistical analysis

Data exclusion

Of the 91 children that were asked to play the story game, seven children were excluded from the dataset because of technical game problems during the intervention period (bugs, explained in Multimedia Appendix 1). Two additional children were excluded because they never started the general digital gaming intervention properly (i.e., they played less than 10% of GG-FL/AC and the story game). Hence, the final number of children for whom useful story game data were available was 82. Data visualization and statistical analyses were conducted in RStudio [70,71].

Enjoyment analysis

Story game enjoyment was addressed in three ways. First, post-intervention child and parental questionnaire outcomes were analyzed by visualizing the relative frequencies per response category of each question. Second, two members of the research group independently evaluated the GameFlow criteria by assigning a state of either fulfilled, partly fulfilled, or not fulfilled. Then, after reaching a consensus, all criteria received a final single state. Third, in-game enjoyment-related data were evaluated, such as general story appreciation, measured as the proportion of given likes and dislikes per story.

Feasibility analysis

In order to gain insights into the feasibility of the intervention, we first defined the general question response accuracy, which indicated the difficulty level and thus the feasibility of the story listening part of the game. It was computed as the proportion of correct and incorrect responses per question. Second, we visualized the attrition rate, which determined how many children completed the intervention and at what point eventual drop-outs occurred. Third, we calculated the individuals' final game exposure and training duration. Final game exposure was computed as the ratio (%) of actually played sessions and the total available sessions (i.e., 72). Individuals' corresponding final training duration represented the amount of days the child played, starting from the day of the first story session until the day on which the participant played for the last time, irrespective of game completion. The measure of final training duration was particularly informative on the feasibility of the training intensity. By analyzing players' final game exposure and training duration, we categorized players into different gaming profiles based on intervention completion (complete/incomplete) and schedule compliance (compliant/non-compliant). Categorizing participants based on these criteria provides information on the possible need to adjust the training schedule and intervention duration to increase feasibility in future studies.

Four additional non-parametric median-based Theil-Sen regressions were performed using the 'mblm' package in R [72] to predict either individuals' final game exposure or final training duration, either based on initial individual story appreciation or question response accuracy. These analyses provide information on how to keep players engaged from the start and prevent them from dropping out or slowing down. The Theil-Sen technique allows for a robust line fitting as its estimation is obtained by calculating the median of the slopes of all possible pairs of datapoints [73]. This non-parametric regression technique was opted for given the relatively small sample size and violated assumptions to perform ordinary least square regressions. Initial individual story appreciation and individual response accuracy at the start of the intervention were defined as a player's mean story appreciation and mean question response accuracy of the first two game phases respectively (i.e., mean ratings of the first 12 stories and mean response accuracy for the first 24 questions). As for predicting game exposure based on initial story appreciation or initial individual question response accuracy, the analyses were restricted to the sample of players that did not completely finish the story game (n = 32) as including participants who completed the full intervention would render a ceiling effect in the results. As for predicting training duration based on initial story appreciation or individual response accuracy, we excluded eight participants who played less than 80% of the total game content, since their training duration in days did not represent a reliable intervention trajectory. As such, training duration was predicted based on data of 74 participants.

Analyzing growth in language comprehension

The third and last part of the result section tackles a possible growth in language comprehension. We considered the mean question response accuracy (%) per game phase as an indication of language comprehension. However, we acknowledge that stories and questions did not belong to a validated language instrument and we did not directly test the

stability of the difficulty of the questions and stories. Yet, we were obliged to select measures beyond existing language instruments given the long intervention period. Each game phase contained 12 content-related questions. Using the 'lmerTest' package in R [74], we performed a multilevel linear growth model in the sample that finished at least 80% of the game (n = 74) with game phase as a within-subjects variable and baseline listening comprehension, morphological awareness, and receptive vocabulary as between-subjects covariates. Game phase was coded such that the fixed intercept of the model represented the mean question response accuracy of the first game phase if a score of 0 would have been obtained on all three baseline language covariates. We also modelled a randomly varying intercept and slope across individuals and added three additional phase*baseline language interaction terms (one per baseline language covariate). As such, we could test the hypothesis whether mean question response accuracy, our indication of language comprehension, improved over game phase, when controlling for each baseline language skill, but also whether children with lower baseline language skills showed larger growing potentials concerning mean question response accuracy.

Results

Enjoyment

Post-intervention questionnaires

Post-intervention child questionnaire data revealed that 61 participants (74%) liked the game very much (see left panel of Figure 4). Moreover, 54 participants (66%) confirmed that they would be willing to redo the training (see right panel of Figure 4). Post-intervention parental questionnaires revealed (1) that the majority of parents observed a high (n = 39, 47%) or relatively high play motivation (n = 32, 39%) in their child (see left panel of Figure 5), (2) that most of the players never (n = 43, 52%) or only sometimes (n = 25, 30%) needed

encouragement to play (see middle panel of Figure 5), and (3) that more than three quarters of the parents (n = 65, 79%) observed a state of sustained attention during gameplay (see right panel of Figure 5).

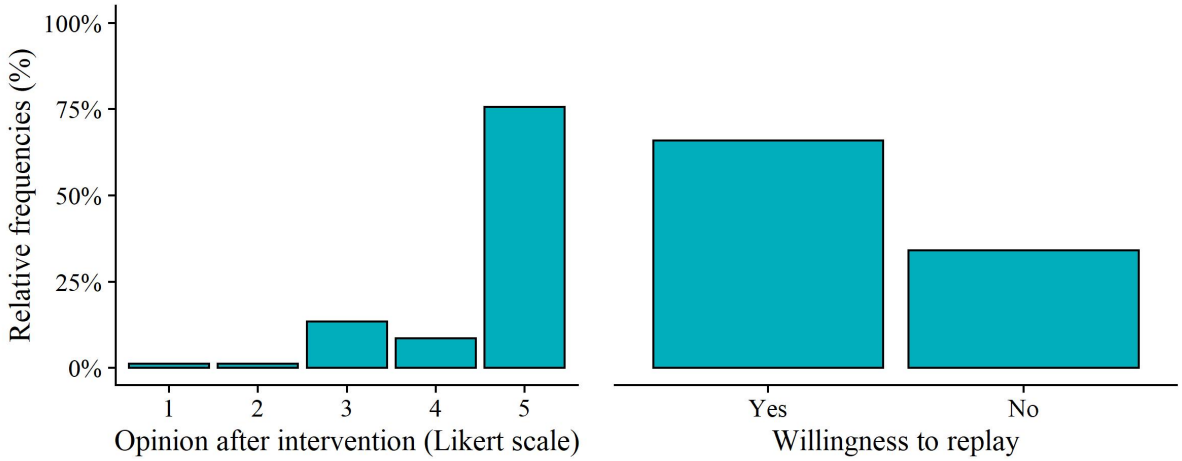


Figure 4. Child questionnaire outcomes. Left panel) Relative frequencies of the post-intervention ratings assigned by the child on a five-point-Likert-scale-based Smiley-o-meter ranging from 1 (very unhappy face) - "I did not like the game at all" to 5 (very happy face) - "I liked the game very much". Right panel) Relative frequencies of the willingness to replay the game.

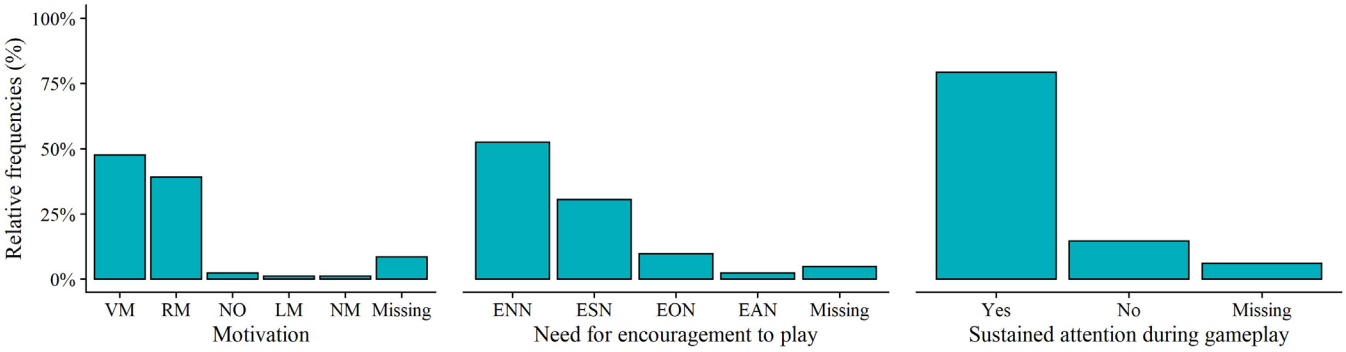


Figure 5. Parental questionnaire outcomes. Left panel) Relative frequencies of the play motivation rated by parents. Middle panel) Relative frequencies of the opinion on

encouragement needed to play. Right panel) Relative frequencies of the opinion on the absence or presence of sustained attention during gameplay.

Note. VM. Very motivated. RM. Relatively motivated. NO. No opinion. LM. Little motivated. NM. Not motivated. ENN. Encouragement never needed. ESN. Encouragement only sometimes needed. EON. Encouragement often needed. EAN. Encouragement always needed.

GameFlow criteria fulfilment

Although the GameFlow model was applied in the design process, Table 1-8 clarifies to what extent game criteria were properly implemented and if our game did, only partly, or did not fulfil each GameFlow criterion [5]. As shown in Table 1-8, the GameFlow-model-based evaluation revealed that only the criteria belonging to the concentration principle were completely fulfilled. As for the other elements, except for the elements of social interaction (as it was intentionally not implemented given the young age of our participants) and challenge, we concluded that most of their corresponding criteria were completely or partly fulfilled. In order to avoid plagiarizing from Sweetser and Wyeth [5], the wording of the descriptions of the GameFlow criteria in Table 1-8 differs from the criteria wording used in their article [5], albeit the meaning of the criteria did of course not change. However, we want to emphasize that we considered the original wordings when developing and evaluating the story game, as we wanted to take into account the importance of applying evaluation tools in the exact same way as how they are validated.

Table 1. Corresponding criteria of the GameFlow element "concentration" and their implementation and fulfilment in the story game

Concentration criteria	Implementation in the story game?	Fulfilled?
Stimuli are provided in different modalities	Stimuli were aurally and visually presented both in the main intervention task (e.g., story audio and accompanying illustrations) as well as in the virtual hub world and avatar customization system (e.g., simple tunes and visual animations) at all times.	Yes
Stimuli are worth	Story texts and corresponding story images were	Yes

attending to	selected based on the target age. The virtual hub world and the avatar customization system provided joyful simple tunes and appeared with flat primary colors and simplified basic models and animations, which are suitable for our target age group [64,65]. Avatars were designed with exaggerated cartoon-like proportions. As such, stimuli seemed appealing, eye-catching, and worth attending to for a long time.	
Games must quickly catch and hold the players' attention and must be able to maintain focus at all times		Yes
There are no irrelevant game tasks on top of the main task	When children were required to listen to the story (i.e., the main learning task), the game did not require performing other irrelevant side-tasks.	Yes
Games should require cognitive workload but should not exceed perceptual and cognitive limits	Players were required to actively listen and stay focused during the whole story listening phase, since they could only collect rewards (coins) when they responded correctly to the content-related questions. However, stories and content-related questions were chosen to correspond to the cognitive limits of five-year old kindergarteners. The virtual hub world and avatar customization system warranted easy-to-use mechanics and simple colorful animations so that players did not become "lost in the game world" or overstimulated.	Yes
There are no distractors during the game tasks	Tasks in the main intervention and avatar customization phase were clearly defined, such that, when performing the tasks (e.g., actively listening to the story, responding to the questions, or spending coins in the avatar customization system), players were not able to perform any other tasks and were not distracted by irrelevant stimuli.	Yes

Table 2. Corresponding criteria of the GameFlow element "challenge" and their implementation and fulfilment in the story game

Challenge criteria	Implementation in the story game?	Fulfilled?
Challenges are adjusted to players' skills	Since the research group selected age-appropriate stories and invented literal questions that were supposed to match the cognitive skills of five-year old kindergarten children and since the virtual hub world and avatar customization system appeared self-explanatory without any form of written language, we concluded that the game tasks were adapted to the skills of the target pre-reading age group.	Yes
Challenges are player-adjusted	The selection and the order of the stories along with their corresponding content-related questions were fixed and the same for all players. Moreover, although we have no direct proof of the stability of the difficulty of the stories and questions, the game was originally not designed to increase in difficulty. In that sense, we concluded that the challenge levels were not individually adapted and did not increase along with	No
Challenge levels increase to improve the players' skills		No

	progress in player skills.	
The game provides new challenges at appropriate times	Although not directly tested, we assumed that the difficulty level of the stories, the questions and other tasks remained relatively stable throughout the game. However, book series and storytellers switched every four sessions. This required adaptability and possibly formed a new challenge for some (but not all) players.	Partly

Table 3. Corresponding criteria of the GameFlow element "player skills" and their implementation and fulfilment in the story game

Player skills criteria	Implementation in the story game?	Fulfilled?
The game is playable without an instructive manual	Although the research group provided a short manual for the players' parents/caregivers, the game was designed in a way that five-year old children could pass the first and following sessions without any form of explanation. Game mechanics were easy-to-use, transfers from one game mode to another occurred automatically, and since the target group was considered pre-reading, there was no form of written text implemented in the game.	Yes
Learning the game is not boring	A learning phase or tutorial game was not implemented, since game sessions were self-explanatory and players learned to play by doing. When opening the application on the tablet for the first time, players gained enough information on how to progress. Hence, we believed that these criteria did not necessarily need to be implemented.	No
Absorbing tutorial games should teach players how to play the game		No
Online help is provided for players in need	The research group did not provide an online help-service tool. However, parents/caregivers could contact the research group via e-mail or telephone anytime in case of technical/motivational problems. If necessary, a member of the research group visited the players at home to fix possible technical/player-related problems.	Partly
Skill progress occurs gradually at an appropriate pace	Although not directly tested, we assume that the difficulty level of the questions and other tasks remained relatively stable throughout the game and that they were adapted to the target age. The game did primarily not intend to specifically improve language comprehension (measured based on mean question response accuracy per game phase) of the target group. However, it is possible that for some of the players, mean question response accuracy would gradually increase with higher story game exposure. This will be explored in the current study.	Partly
Skill effort and development is rewarded	Skill effort was rewarded by means of stars. After finishing a session, regardless of whether the responses to the content-related questions were correct or not, all players earned a star, which appeared on a stage in the virtual hub world and indicated the players' progress in	Yes

	the game. Skill development was rewarded with coins, such that, the more correct responses were given, the more coins were earned.	
Game mechanics and interface are simple and easy to apply	All actions required the use of touch screen, since the game was tablet-based. Nowadays, in Western society, the majority of five-year olds are familiar with these devices. Furthermore, players only had a limited set of actions, as advanced game options were locked with a password. These limited actions (e.g., stopping, starting, or continuing the story game, selecting a response for the questions, buying accessories in the avatar customization system) were all self-explanatory (e.g., selecting the correct response or a desired accessory by touch screen) or assigned with clear symbols on the screen (red arrow to stop the game, large green play-symbol to start or continue the story recording).	Yes

Table 4. Corresponding criteria of the GameFlow element "control" and their implementation and fulfilment in the story game

Control criteria	Implementation in the story game?	Fulfilled?
A sense of control over the game avatar, its movements and interactions	During the main intervention task, it was not possible to exercise control over the game avatar. Yet, following the main intervention task, players were able to customize their avatar or buy new avatars in the avatar customization system, all of which or whom appeared later on the platforms in the virtual hub world. However, since the avatars automatically jumped to the next stage when finishing a session, their movement and interaction control was limited.	Partly
A sense of control over the game interface and device	Since five year old children are relatively familiar with tablets nowadays, we believed that the tablet was a suitable device for the intervention program. Moreover, individual player profiles were set in advance on each tablet, and the main menu contained no other option than starting or stopping the game, making the game interface intuitive and self-explanatory.	Yes
A sense of control over the game mechanics (saving, stopping, etc.)	Players were able to start and stop the game anytime. However, when interrupting the game during story listening, the progress of the ongoing session was not registered or saved and players should restart the whole session.	Partly
Errors that will harm the game (e.g., bugs) should not occur and if they occur, support must be available	The game was piloted many times on members of the research group and on children with a similar age as the target group. Although the game appeared bug-free at the end of the pilot studies, some players in the actual intervention study (not included in the final data analysis) experienced bugs causing a total crash of the game and loss of in-game data. In those cases, a member of the research group provided support as soon as	Partly

	possible by reinstalling the game at the homes of the players.	
A sense of control over the game world (possibility to shape the game world)	Following the main intervention task and the time spent in the avatar customization system to change the game avatar, the newly dressed/new game avatar appeared in the virtual hub world on one of the stages. In that sense, players exercised control over altering their game world. However, the background animations and models in the hub world were preprogrammed and could not be altered.	Partly
A sense of control over the game actions and strategies	As the order and goal of the game along with the framed intervention task was fixed for all players, it was not possible to make use of different strategies and actions to reach the end of the game.	No

Table 5. Corresponding criteria of the GameFlow element "clear goals" and their implementation and fulfilment in the story game

Clear goals criteria	Implementation in the story game?	Fulfilled?
Final goal is clear from the start	Prior to the very first session, players viewed the virtual hub world from a "drone's perspective" and flew over all 72 stages which their game avatar had to pass to finish the game. That way, the player was immediately informed on the total amount of sessions. However, since the drone cut-scene was only presented at the start without any form of explanation, the final goal (e.g., earning 72 stars) was not clearly emphasized during the entire intervention period.	Partly
Intermediate goals are clear	After playing a first session, it was clear for the player what should be reached during one game session.	Yes

Table 6. Corresponding criteria of the GameFlow element "feedback" and their implementation and fulfilment in the story game

Feedback criteria	Implementation in the story game?	Fulfilled?
Feedback on progress	After every session, feedback on the game progress was visually provided by the game avatar who jumped from one stage to the next in the virtual hub world. A star appeared on the platform that was just left. When the avatar reached the 72nd platform, all game sessions were finished.	Yes
Immediate feedback on actions	A star was given to all players as immediate feedback on the session completion. When responding correctly to a content-related question, the player received a coin, which was immediately presented after assigning the correct answer with animations and sound effects. Moreover, newly bought accessories or avatars immediately appeared in the virtual hub world.	Yes
Status or score is presented at all times	The amount of collected stars was only visible in the main menu. The amount of earned coins was visible in	No

	the avatar customization system or could be considered when clicking on a little wallet that was visible in the virtual hub world. Thus, during the main intervention task, no score status was presented since we wanted to prevent players from being distracted by irrelevant stimuli.	

Table 7. Corresponding criteria of the GameFlow element "immersion" and their implementation and fulfilment in the story game

Immersion criteria	Implementation in the story game?	Fulfilled?
Decreased awareness of surroundings when playing the game	Although not measured and not measureable with questionnaires in our young target age-group, Figures 4 and 5 showed that the majority of the players liked the game very much, were motivated to play, and would be willing to play again. Moreover, the majority of the parents/caregivers indicated a state of full focus (as shown in the right panel of Figure 5) when their child played the game, suggesting that some form of immersion was present in the majority of the players.	N.M. [a]
Decreased self-awareness and less everyday worries when playing the game		N.M. [a]
A changed sense of time		N.M. [a]
Emotional involvement in the game		N.M. [a]
Visceral involvement in the game		N.M. [a]

[a] N.M. Not measured.

Table 8. Corresponding criteria of the GameFlow element "social interaction" and their implementation and fulfilment in the story game

Social interaction criteria	Implementation in the story game?	Fulfilled?
Competition between different players	The game did not intend to create competition between different players as participants in the current study were recruited anonymously and independently from each other.	No
Social interaction between different players (e.g., chat services)	Not applicable for our target age group of five-year old children.	No
Available social communities about the game		No

In-game data related to game enjoyment

Figure 6 visualizes the general story appreciation, calculated as the proportion of available likes and dislikes per story. Each story was liked by more than 75% of the listeners

for whom data were available (i.e., more than 62 of the 82 listeners), indicating that the stories were enjoyable.

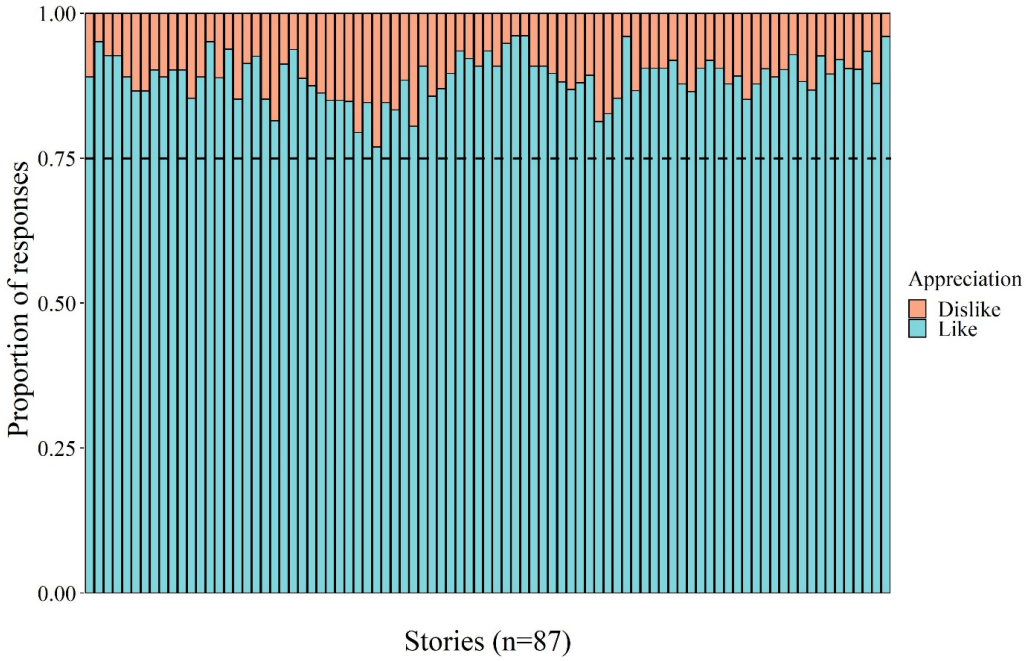


Figure 6. Overview of the general story appreciation.

Note. Each vertical bar represents one story and stories are ordered based on occurrence in the game.

Feasibility

General question response accuracy

Figure 7 visualizes the general question response accuracy, calculated as the proportion of correct and incorrect responses per question. All but 11 questions were answered correctly by 75% of the listeners for whom data were available (e.g., 62 of the 82 listeners), suggesting a rather stable difficulty level of the questions.

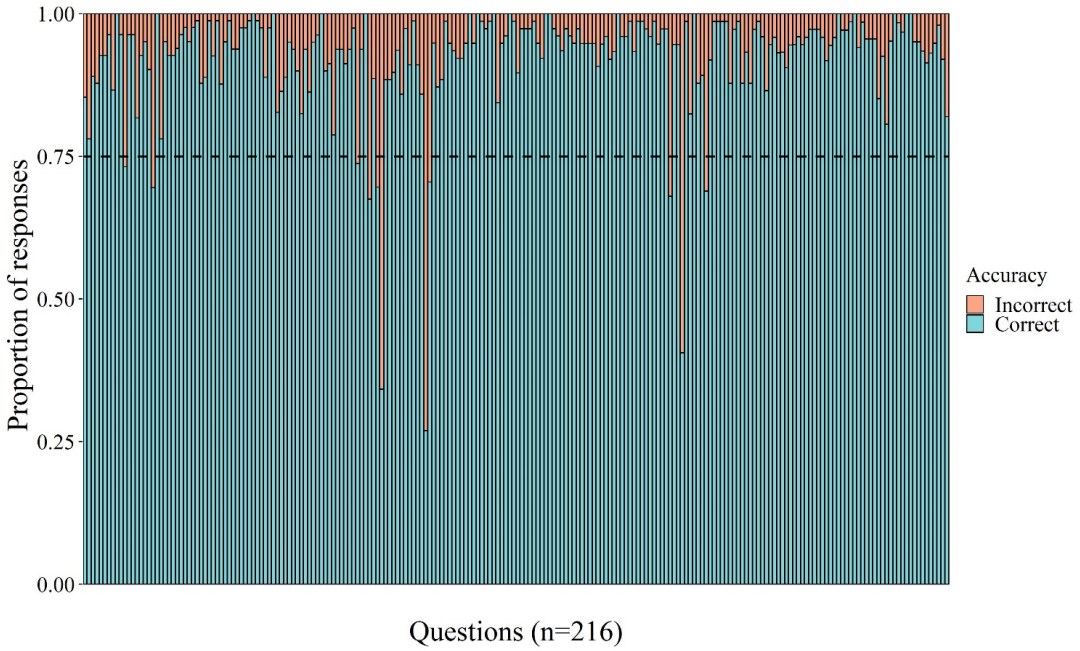


Figure 7. Overview of the general question response accuracy.

Note. Each vertical bar represents one question and questions are ordered based on occurrence in the game.

Attrition rate

Figure 8 represents the attrition rate and shows the proportion of active participants and drop-outs throughout the entire story game progress. Of the 82 players, 90% (n = 74) listened to 80% of the stories (~ 57th session of the total 72 story sessions) and 61% (n = 50) managed to finish the game completely. Hence, 39% of the players (n = 32) dropped out at some point during the 12-week intervention. Figure 8 also shows that the first player dropped out after the 12th story, which corresponds to the start of the third game phase. From that point onwards, drop-out proportion stayed relatively stable until the 80% completion-point, after which it gradually increased towards the end of the game.

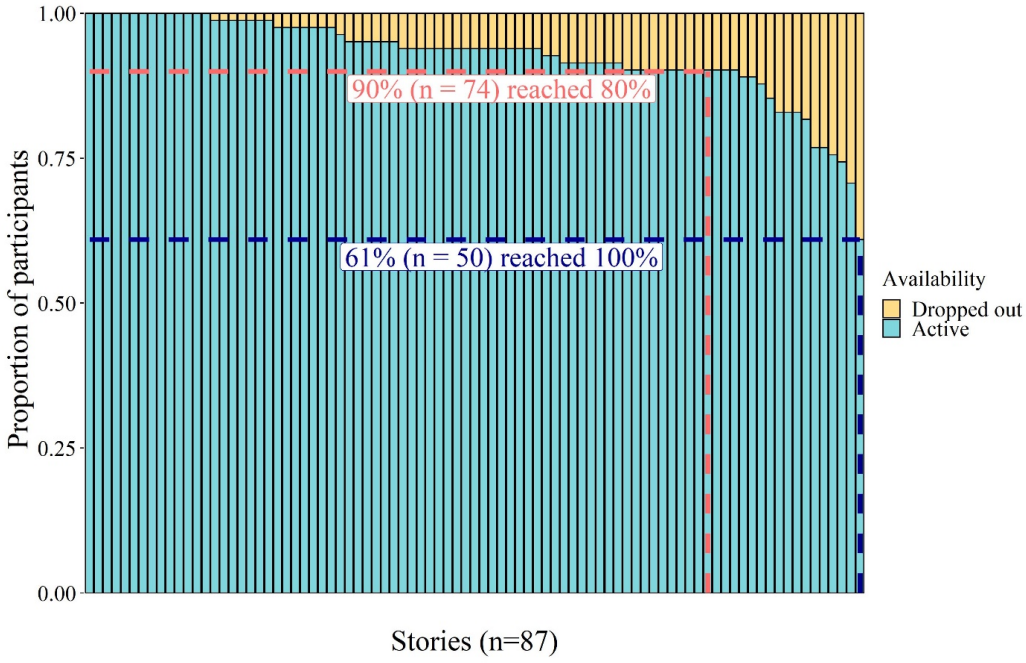


Figure 8. Attrition rate and drop-out occurrence throughout the story game.

Note. Each vertical bar represents one story and stories are ordered based on occurrence in the game.

Final game exposure, training duration and gaming profiles

Figure 9 demonstrates individual story game trajectories, the individuals' final game exposure and corresponding training duration, and their overall distributions. Apart from the visible finding that a few participants had a low training duration simply because they dropped out early, Figure 9 also shows a large variation in training duration in children that nearly or completely finished the story game. Based on the categorization criteria (complete/incomplete and compliant/non-compliant), we established three different gaming profiles: (1) complete and compliant players (n = 18) (upper panel of Figure 10), (2) complete, but non-compliant players (n = 32) (middle panel of Figure 10), and (3) incomplete and non-compliant players (n = 32) (lower panel of Figure 10). Note that we did not observe any players who perfectly followed the training schedule, but dropped out at some point.

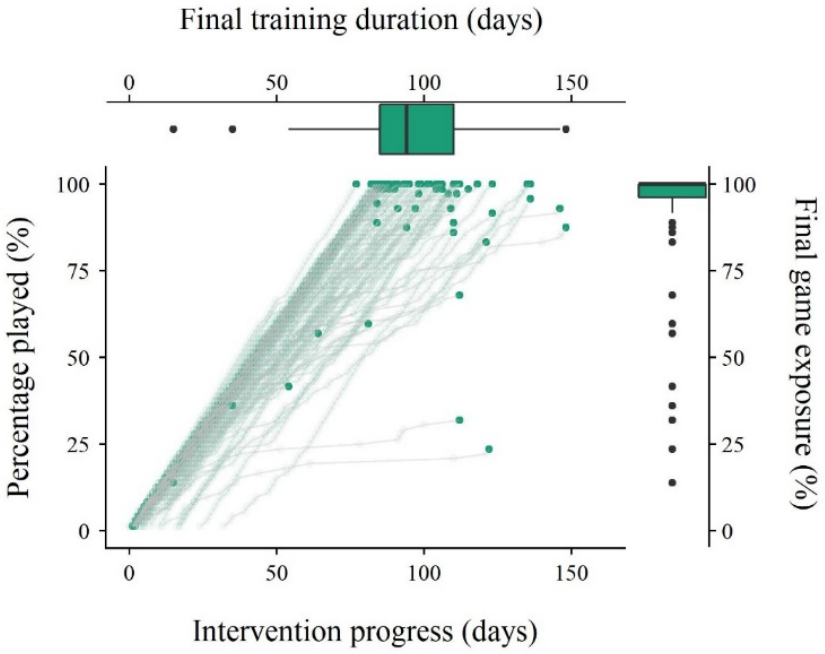


Figure 9. Individuals' game exposure and training duration and their overall distributions.

Note. Bold dots represent final training duration and final game exposure.

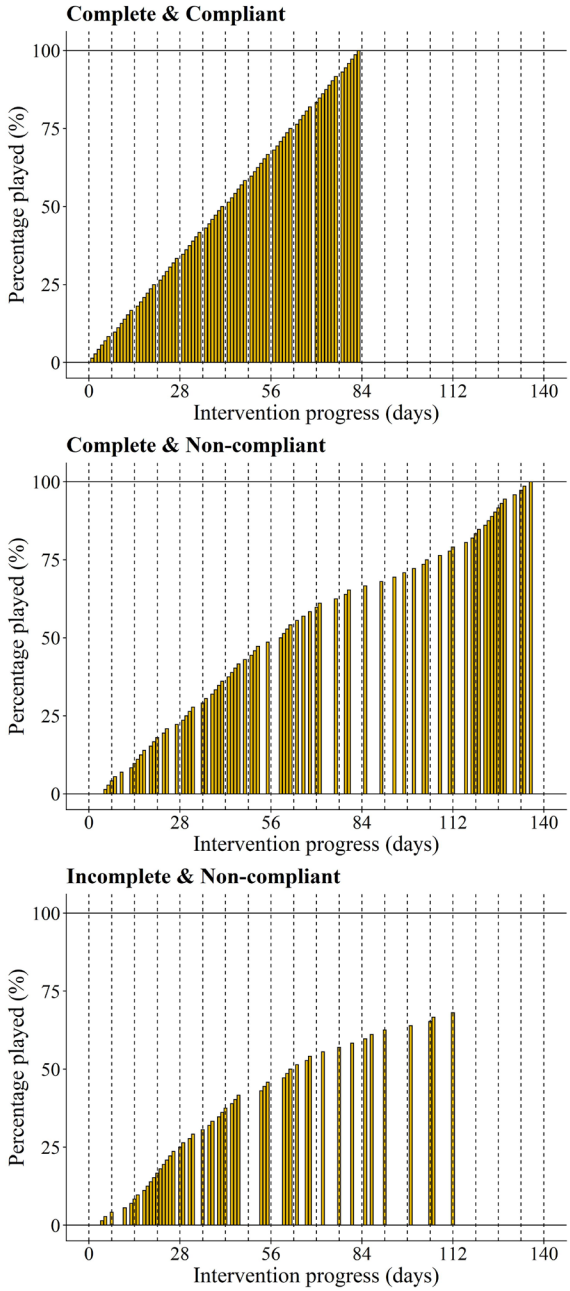


Figure 10. Examples of the different gaming profiles of the story game. Upper panel) Complete and compliant player. Middle panel) Complete and non-compliant player. Lower panel) Incomplete and non-compliant player.

Note. Dotted lines represent the number of intervention weeks.

The Theil-Sen regression analysis revealed that the mean story appreciation of the first two game phases was not significantly predictive of final game exposure among the 32 children who dropped out at some point ($\beta = 0.15$, $P = .08$). The left panel of Figure 11 indeed

shows that the majority of children who quit at some point still assigned positive ratings to all the stories of the first two phases. In contrast, the Theil-Sen regression analysis revealed that the mean question response accuracy of the first two game phases significantly predicted final game exposure, such that a lower accuracy at the start accelerated the drop-out point (β = 0.35, $P < .001$) (see right panel of Figure 11).

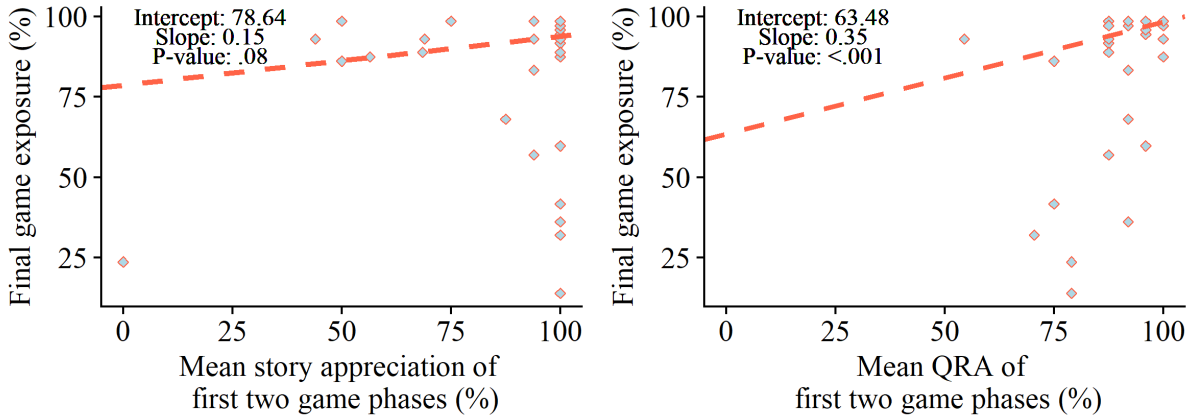


Figure 11. Theil-Sen regression outcomes. Left panel) Predictive relationship between initial mean story appreciation and final game exposure. Right panel) Predictive relationship between initial mean question response accuracy and final game exposure.

Note. QRA. Question response accuracy.

Story appreciation of the first two game phases significantly predicted final training duration in a negative way among those 74 children who finished 80% of the game (β = -0.16, P = .003) (left panel of Figure 12). These results indicate that players who liked the stories at the start were more likely to play the game at a higher rate. Mean question response accuracy at the start did not significantly predict final training duration (β = 0.00, P = .68) among the children who finished or almost completely finished the story game (right panel of Figure 12).

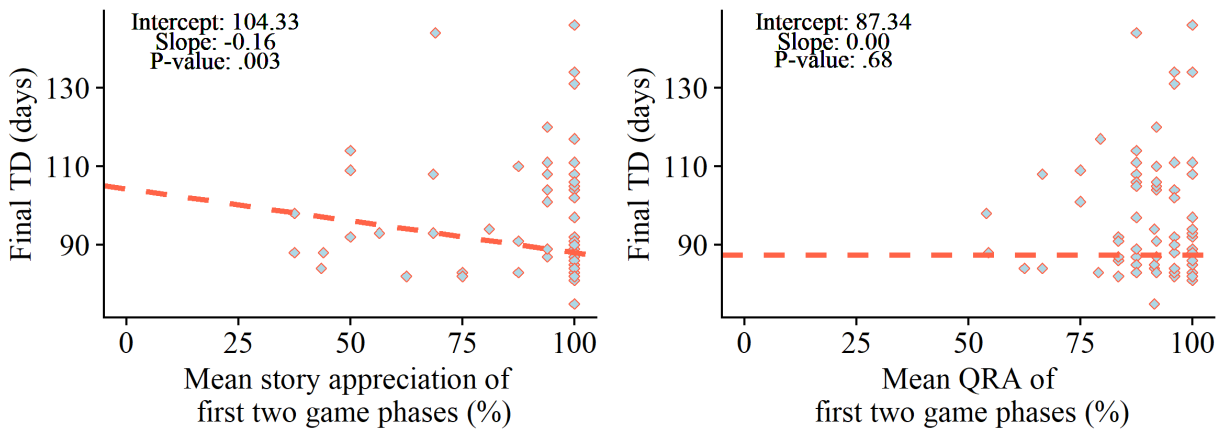


Figure 12: Theil-Sen regression outcomes. Left panel) Predictive relationship between initial mean story appreciation and final training duration. Right panel) Predictive relationship between initial mean question response accuracy and final training duration.

Note. TD. Training duration. QRA. Question response accuracy.

Growth in language comprehension

Figure 13 presents a distribution of the mean question response accuracy of the first game phase and shows overall high initial accuracy scores, suggesting a small growing potential in a large proportion of the children. Model outcomes for the linear mixed effects model are presented in Table 9. Game phase significantly predicted mean question response accuracy in a positive direction (β = 1.44, SE_β = 0.49, $t_{120.12}$ = 2.974, P = .004). Moreover, there was a positive predictive relationship of baseline listening comprehension (β = 1.56, SE_β = 0.49, $t_{71.91}$ = 3.168, P = .002) and receptive vocabulary (β = 0.16, SE_β = 0.06, $t_{72.27}$ = 2.64, P = .01) on the mean question response accuracy. Remarkably, results also revealed a significant interaction of game phase and listening comprehension, such that children with lower baseline listening comprehension underwent a significantly larger growth in mean question response accuracy throughout the game (β = -0.08, SE_β = 0.04, $t_{119.83}$ = -2.03, P = .04).

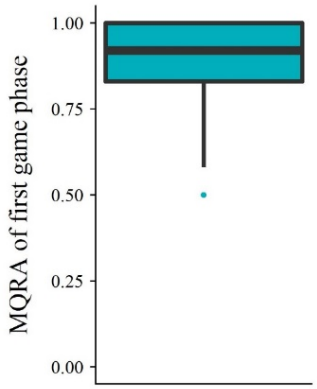


Figure 13. Distribution of the mean question response accuracy of the first game phase.

Note. MQRA. Mean question response accuracy.

Table 9. Results of the linear mixed model analysis.

Model term	Estimate	SE Estimate [a]	$t_{degrees\ of\ freedom}$	*P*-value
Intercept	63.38	6.49	$t_{71.92} = 9.77$	<.001*
Game phase	1.44	0.49	$t_{120.12} = 2.97$	.004
Listening comprehension	1.56	0.49	$t_{71.91} = 3.17$	.002
Receptive vocabulary	0.16	0.06	$t_{72.27} = 2.64$	.01
Morphological awareness	0.23	0.32	$t_{72.19} = 0.73$	.47
Game phase * Listening comprehension	-0.08	0.04	$t_{119.83} = -2.03$	.04
Game phase * Morphological awareness	-0.02	0.02	$t_{123.13} = -0.60$	.55
Game phase * Receptive vocabulary	-0.01	0.01	$t_{123.71} = -1.27$	.21

[a] SE: Standard Error.

* Significant at the .05-level

Discussion

Principal findings

The current study addressed enjoyment and feasibility of a story listening game and preliminarily assessed a possible growth in language comprehension. Principal findings on the

game enjoyment, feasibility, and the impact of the game on language comprehension will be disclosed below.

As for the enjoyment evaluation, child and parental questionnaires and in-game enjoyment-related data (e.g., story appreciation) point to a highly positive game experience. Indeed, questionnaires revealed a high game enjoyment, motivation, and sustained attention during gameplay in most of the participants. The in-game data indicative of game enjoyment also showed that each story was appreciated by minimally three-quarters of the listeners, indicating that the stories fitted the interests of five-year old children at cognitive risk for dyslexia and that most of the participants experienced the story game as enjoyable. Nevertheless, although a considerable amount of GameFlow criteria were fully and partly implemented in the story game, probably benefitting game enjoyment experience to a large extent, the GameFlow-model based evaluation also implied room for optimization of the game design. Indeed, although the GameFlow model [5] was taken into account during the design process of the story game, the concentration principle remained the only principle for which all criteria were completely fulfilled.

Concerning the feasibility of the intervention in terms of difficulty, results suggest that the game was adjusted to the cognitive capacities of five year old children with an elevated risk for dyslexia, since almost all questions were answered correctly by minimally three quarters of the listeners. This was not surprising, since we intended to develop a game, which was not too easy, nor too difficult to ascertain that players were cognitively able to gather many coins, but only if they actively listened to the stories. This state of active listening was confirmed by the general question response accuracy results. The fact that only 11 questions were not answered correctly by 75% of the listeners also suggests a rather stable difficulty level of the questions. As for feasibility in terms of intervention completion and duration, outcomes indicate a positive evaluation on the one hand, as the majority of the participants

were able to fully or nearly complete the story game, but a negative evaluation on the other hand, since we observed a large variation in training duration (e.g., days needed to nearly/completely finish the intervention). These findings thus mainly raise doubts on the feasibility of the training intensity, since many participants could finish, but were simply not able to follow the rather highly concentrated advised schedule of six days per week during 12 weeks. This was also confirmed by the gaming profile outcomes, which revealed that only 18 participants belonged to the "complete and compliant group" and completed the game according to the postulated schedule compared to 32 participants who managed to finish, but spaced out the sessions over a larger time period (e.g., the complete and non-compliant group). As for predicting individuals' final training duration and game exposure, we established a significantly negative relationship between initial mean story appreciation and final training duration and a positive relationship between initial mean question response accuracy and final game exposure. These findings indicate that initial compromised story appreciation decelerated the advised training scheme and that a higher response accuracy at the start increased the amount of played sessions.

As for the growth in language comprehension, we established an increase in mean question response accuracy along with game phase, baseline vocabulary, and baseline listening comprehension. Yet, most strikingly, children with lower listening ability scores at the start of the intervention period made significantly more progress in terms of question response accuracy than children with higher pre-intervention listening ability scores. Overall, these results carefully suggest a potential of the game to train language comprehension with larger gains in children who already exhibit lower language levels at the start.

Limitations and comparisons to prior work

Enjoyment

The game enjoyment-related results give rise to two important discussion points. First, caution is required when interpreting the positive questionnaire outcomes. On the one hand, the phenomenon of social desirability, i.e., the tendency to provide answers which correspond to widely accepted social norms, could have biased the responses of both children and parents [75]. On the other hand, the short attention span and underdeveloped cognitive capacities in young children could have resulted in extremely positive, but less reliable responses on the Smiley-o-meter. Indeed, in line with our study, Zaman and colleagues [76] established an overrepresentation of positive responses when applying the Smiley-o-meter in young children below the age of seven. Accordingly, in another study, authors stated that children equal to or below the age of seven do not yet grasp the principles of a visual analogue scale, from which the Smiley-o-meter is an example [77]. As put forward by Zaman and colleagues [76], the search for how to optimally measure enjoyment of technology in preschoolers is still ongoing. The second discussion point related to enjoyment regards the GameFlow-based evaluation. More specifically, we acknowledge that the agreement of the developed game with the GameFlow model contains a form of subjectivity, as the evaluation was performed by two members of the research group who were involved in the design process itself. Irrespective of this limitation, the fully and partly implemented GameFlow criteria embedded in the current game were presumably satisfactory enough to attain a certain level of enjoyment. This assumed enjoyment level was also supported by the cautiously interpreted questionnaire and in-game enjoyment-related outcomes, which indicated game appreciation in the majority of the players. Optimizing the partly or non-fulfilled criteria (e.g., fixing the remaining bugs, clarifying the end goals) could increase the level of enjoyment even more. To deal with the

subjectivity, a future study could let adults who were not involved in the design process (e.g., parents of playing children) rate the game based on the 0-5 rating procedure.

Feasibility

The current paragraph will discuss three relevant discussion points related to the feasibility outcomes. First, although the majority of the participants (n = 50, 61%) completed the intervention, we still acknowledge that 32 participants (39%) dropped out at some point, yet mostly in the last game phases. This slightly corresponds to another home-based preventive intervention study in Dutch children at risk for dyslexia, which reported an attrition rate of 34% [78]. Unfortunately, the attrition rate of the current study still exceeded the desirable benchmark of 30% [79] and further research on why participants tended to drop out at some point is necessary. In the study of Justice and colleagues [80] for example, who investigated the feasibility, efficacy, and social validity of a 12-week home-based shared storybook reading intervention in preschoolers with language difficulties, authors compared completers with non-completers on a variety of child- and parent-specific characteristics and established risk factors of drop-out, such as lower maternal age and lower parental educational levels. Distinguishing completers from non-completers based on intrinsic and extrinsic features did not belong to the scope of the current study, albeit the research methods applied in the study of Justice and colleagues [80] offer possibilities for future research which could be of added value to optimize feasibility of the current intervention program. Irrespective of the rather high attrition rate, we still established a completion of 80% (~ 57th session of the total 72 story sessions) in 90% of the participants (n = 74), suggesting that shortening the intervention with approximately 15 sessions, which almost equals four weeks in the current study, would result in a lower drop-out occurrence. Yet, an important sidenote relates to the fact that participants in the current study combined the intervention with another tablet game (GG-FL or AC games). This could also have impacted general motivation and

perseverance. The second discussion point related to feasibility bears upon the variation in training intensity among participants. Since treatment fidelity in terms of quality and quantity has been found to predict learning outcomes in previous studies [78,81], the variation in training intensity cannot be neglected when investigating the educational impact of interest in a future study. However, in their reading intervention study, Katzir and colleagues [82] found larger gains on sight word efficiency in a group of grade 1-3 children at risk for reading difficulties who received a 44-hour fluency-based therapy over a nine month period compared to children who followed the exact same intervention within a more intensive period of two months. Hence, albeit the need for statistical confirmation in a future study, spacing out the story game intervention over time in the current study might have even benefitted the expected educational learning outcomes. The last feasibility-related discussion point is linked to the prediction of final game exposure and training duration based on story appreciation and question response accuracy at the start of the intervention. The outcomes of these predictive analyses point to limitations in the current game design, but give rise to suggestions for optimizing intervention feasibility in future studies. For example, given the importance of initial story appreciation to stick at the advised training intensity, a possible suggestion involves offering the story book series in the preferred order of the participant. More specifically, providing a main catalog menu in which players could choose which story series occur first might increase the story appreciation at the beginning and as a result, the engagement to follow the advised schedule. Moreover, given the role of initial question response accuracy in drop-out occurrence, offering player-adapted questions based on an individual's language knowledge and cognitive capacities, which then increase in difficulty, instead of fixed predefined questions for all participants, might prevent participants to withdraw from the study. A more interactive approach in which players could request and receive explanations of possible difficult words might also increase the chance to already

successfully respond to the questions from the start, lowering the chances of early drop-out. However, a disadvantage of these proposed adjustments to prevent attrition relates to the fact that growth of in-game question response accuracy data then becomes less easily interpretable at the group level, since participants would follow an individually adapted trajectory. However, large-scale adaptability has been widely considered as one of the strengths of serious digital gaming interventions [4]. Furthermore, according to our GameFlow-based evaluation in its current form, the criterion of so-called "player-adjusted challenges" belonging to the overall challenge principle is lacking in the current story game. Although this principle was beyond the scope of the story game, its implementation could possibly increase player enjoyment and decrease attrition rate [5].

Growth in language comprehension

The results related to growth in language comprehension must be interpreted with extreme caution and several important limitations should be elucidated. The first and most important limitation relates to the research design. In fact, the actual impact of the game on language comprehension can only be solidly established by conducting a randomized controlled trial (RCT), which was not performed in the current study. Hence, although we found a larger growth in mean question response accuracy in children with lower listening comprehension, the current research design does actually not allow us to draw univocal conclusions on the potential of the game to train language comprehension. A future RCT study, preferably including (1) a group that is purely playing the story game without combining it with other interventions such as GG-FL or active control games, (2) a no-intervention control group, and (3) a control group receiving an alternative placebo treatment which does not specifically train language skills [83], is thus of crucial relevance to further disentangle the gaming effects on language comprehension. A second important limitation relates to the concept of content validity. We considered the in-game questions as a measure

of language comprehension. However, albeit the selection of age-appropriate stories based on library visits and the construction of the questions based on the vocabulary of the story content, the stories and questions did not belong to a validated comprehension test instrument, casting doubts on the certainty that we truly measured language comprehension in our participants [84]. In addition, we are unsure whether all questions and stories implemented in the game were of equal difficulty, although we found relatively stable response distributions in the majority of the questions. Hence, changes in performance might not be due to the intervention effects, but simply to the varying difficulty of the questions and stories. Conducting pre- and post-intervention measurements of independent language comprehension tests within an RCT design with appropriate control groups (as aforementioned), would eliminate this concern. The third limitation discusses the distribution of the mean question response accuracy in the first game phase, which already showed a tendency towards ceiling. Hence, children with high accuracy scores from the start, related to a higher baseline vocabulary and listening comprehension, did not have the potential to increase any further during the intervention, influencing the interpretation of the game phase-listening comprehension interaction. The high accuracy scores at the start are presumably attributed to the fact that none of the participants in the current study experienced language difficulties. Indeed, originally, the story game was not developed to train language comprehension within the framework of the current research project, but rather to investigate the potential of envelope enhanced speech in children with a pure cognitive risk for dyslexia. Another factor, possibly influencing the high initial accuracy scores, relates to the type of posed questions. Questions intentionally only involved identifying or recalling items that were explicitly mentioned in the story text and thus examined language comprehension only at the literal language level [85]. Hence, questions at the inferential level, which for example require implying emotions, predictions, and connections between information within the text or with

the child's own experiences [86], were not part of the story game in its current form. For preschoolers, inferential questions are certainly more difficult to answer than literal ones [85]. Yet, given the scope of the story game within the current research project, i.e., exposing children at risk for dyslexia in a joyful way to either envelope enhanced or non-enhanced speech, questions were not supposed to be or become difficult, but only served as a way to (1) ensure active listening and (2) to ascertain that players earned as much as coins as possible, increasing the chance to maintain the intervention for 12 weeks. Based on these aforementioned limitations, the preliminary and cautiously interpreted results concerning a possible growth in language comprehension give rise to three important ideas for story game improvement if it would actually be used to train language comprehension at some point in populations with specific problems in this domain (e.g., children with low SES, bilingual children, children with DLD, with a hearing impairment, or dyslexia). First, there is a clear need to also include inferential questions, since Tarvainen and colleagues [18] emphasized the importance of inferential language training to foster language comprehension in preschoolers. Second, apart from optimizing the research design by conducting RCTs with adequate intervention and control groups, it is also of crucial relevance to include test batteries, which assess comprehension both at the inferential and literal level. The currently used CELF-4-NL [66] and the Peabody Picture Vocabulary-III-NL tasks [67] mainly focus on language comprehension at the literal level [85], emphasizing the need to add inferential language comprehension test batteries to the research protocol. Last, researchers must take into account and tackle the possibility that children with lower language levels might show less motivation in storybook interventions than their typically developing peers [87].

General conclusions

Overall, our newly developed story game generally seemed feasible and enjoyable for five-year old children with an elevated cognitive risk for developmental dyslexia. Moreover,

our results preliminarily, but carefully point to a potential of the game to train language comprehension. Hence, we hope to apply the story game in follow-up studies with an optimized study design in children who, for whatever reason, exhibit actual language comprehension difficulties. As such, we will be able to (1) generalize our enjoyment and feasibility findings, and (2) draw solid conclusions on the actual influence of the game on language comprehension.

Acknowledgements

FVB contributed to the data collection and processing, the project administration, and wrote the original draft. ME was involved in the project administration, data collection and processing, and reviewed and edited the draft. WD programmed the tablet game described in the study and mostly contributed to the multimedia appendices. MV, JW, PG, and JV conceptualized and received funding for the project. Moreover, they supervised the project and were involved in the review and editing part of the writing process. This study obtained funding from the Research Council of KU Leuven, Belgium (C14/17/046). Co-author Jolijn Vanderauwera was a postdoctoral researcher and received a personal grant of the Research Foundation Flanders (FWO), Belgium (12T4818N). We would like to thank all participants, their parents, and the involved school workers for their active and intensive participation in the study. We also acknowledge the highly skilled speakers who were involved in the story recording sessions for a considerable amount of their free time. In addition, we are thankful to the colleagues of the DYSCO team at the research unit of Parenting and Special Education and of the Research Group ExpORL. Finally, we would like to acknowledge the Master Students of Educational Sciences at KU Leuven and job students who intensively cooperated in the data collection at schools.

Conflicts of interest

This study was conducted without any conflicts of interest (“none declared”).

Multimedia Appendix

Multimedia Appendix 1: [Technical story game development guide]

Multimedia Appendix 2: [Demonstration movie of the story game]

Multimedia Appendix 3: [Overview of the child and parental categorical questions and response possibilities]

References

1. Montero G, Gomez J. Serious games in special education. A practitioner’s experience review. In: Clua E, Roque L, Lugmayr A, Tuomi P, editors. Entertainment Computing - ICEC 2018. Cham: Springer International Publishing; 2018:397-401. ISBN:9783319994253
2. Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R. Assessment in and of serious games: an overview. Adv Human-Computer Interact. 2013;2013:1-11. doi:10.1155/2013/136864
3. Furió D, González-Gancedo S, Juan MC, Seguí I, Rando N. Evaluation of learning outcomes using an educational iPhone game vs. traditional game. Comput Educ. 2013;64:1-23. doi:10.1016/j.compedu.2012.12.001
4. Dörner R, Göbel S, Effelsberg W, Wiemeyer J. Serious games: foundations, concepts and practice. Cham: Springer International Publishing : Imprint Springer; 2016. ISBN:9783319406114
5. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. ACM Comput Entertain. 2005;3(3):1-24. doi:10.1145/1077246.1077253
6. Ferreira de Almeida JL, dos Santos Machado L. Design requirements for educational serious games with focus on player enjoyment. Entertain Comput. 2021;38:100413. doi:10.1016/j.entcom.2021.100413

7. Read J, Macfarlane S, Casey C. Endurability, engagement and expectations: measuring children's fun. In: Interaction design and children. Eindhoven: Shaker Publishing; 2002:1-23.

8. Rogde K, Hagen ÅM, Melby-Lervåg M, Lervåg A. Protocol: The effect of linguistic comprehension training on language and reading comprehension: a systematic review. Campbell Syst Rev. 2016;12(1):1-50. doi:10.1002/CL2.161

9. Gough PB, Tunmer WE. Decoding, reading, and reading disability. Remedial Spec Educ. 1986;7(1):6-10. doi:10.1177/074193258600700104

10. Morgan L. An exploration of evidence of deficits and theories of therapy for specific-receptive language impairment in children. Dissertation. University of the West of England; 2013. https://uwe-repository.worktribe.com/OutputFile/932141 [accessed Jul 14, 2021]

11. Snow C, Burns M, Griffin P, Griffin MS. Preventing reading difficulties in young children. Washington, DC: National Academy Press; 1998. ISBN:030906418x

12. Muter V, Hulme C, Snowling MJ, Stevenson J. Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. Dev Psychol. 2004;40(5):665-681. doi:10.1037/0012-1649.40.5.665

13. Tarvainen S, Launonen K, Stolt S. Oral language comprehension interventions in school-age children and adolescents with developmental language disorder: a systematic scoping review. Autism Dev Lang Impair. 2021;6:1-24. doi:10.1177/23969415211010423

14. Hagen ÅM, Melby-Lervåg M, Lervåg A. Improving language comprehension in preschool children with language difficulties: a cluster randomized trial. J Child Psychol Psychiatry. 2017;58(10):1132-1140. doi:10.1111/jcpp.12762

15. Elbro C, Dalby M, Maarbjerg S. Language-learning impairments: A 30-year follow-up of language-impaired children with and without psychiatric, neurological and cognitive difficulties. Int J Lang Commun Disord. 2011;46(4):437-448. doi:10.1111/j.1460-6984.2011.00004.x

16. Durkin K, Conti-Ramsden G. Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. Child Dev. 2007;78(5):1441-1457. doi:10.1111/j.1467-8624.2007.01076.x

17. Melby-Lervåg M, Lervåg A, Lyster S-AH, Klem M, Hagtvet B, Hulme C. Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. Psychol Sci. 2012;23(10):1092-1098. doi:10.1177/0956797612443833

18. Tarvainen S, Stolt S, Launonen K. Oral language comprehension interventions in 1–8-year-old children with language disorders or difficulties: a systematic scoping review. Autism Dev Lang Impair. 2020;5:1-24. doi:10.1177/2396941520946999

19. Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T, CATALISE-2 consortium. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study

of problems with language development: terminology. J Child Psychol Psychiatry. 2017;58(10):1068-1080. doi:10.1111/jcpp.12721

20. Clark A, O'Hare A, Watson J, et al. Severe receptive language disorder in childhood-familial aspects and long-term outcomes: results from a Scottish study. Arch Dis Child. 2007;92(7):614-619. doi:10.1136/adc.2006.101758

21. Melby-Lervåg M, Lervåg A. Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. Psychol Bull. 2014;140(2):409-433. doi:10.1037/a0033890

22. Hart B, Risley TR. The early catastrophe: The 30 million word gap by age 3. Am Educ. 2003;27(1):4-9. ISSN:0148432X

23. Walker EA, Sapp C, Dallapiazza M, Spratford M, McCreery RW, Oleson JJ. Language and reading outcomes in fourth-grade children with mild hearing loss compared to age-matched hearing peers. Lang Speech Hear Serv Sch. 2020;51(1):17-28. doi:10.1044/2019_LSHSS-OCHL-19-0015

24. Rispens J, Been P. Subject-verb agreement and phonological processing in developmental dyslexia and specific language impairment (SLI): a closer look. Int J Lang Commun Disord. 2007;42(3):293-305. doi:10.1080/13682820600988777

25. Casalis S, Leuwers C, Hilton H. Syntactic comprehension in reading and listening: a study with French children with dyslexia. J Learn Disabil. 2013;46(3):210-219. doi:10.1177/0022219412449423

26. Ebbels S. Effectiveness of intervention for grammar in school-aged children with primary language impairments: a review of the evidence. Child Lang Teach Ther. 2014;30(1):7-40. doi:10.1177/0265659013512321

27. Kendeou P, van den Broek P, White MJ, Lynch JS. Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills. J Educ Psychol. 2009;101(4):765-778. doi:10.1037/a0015956

28. Foorman BR, Herrera S, Petscher Y, Mitchell A, Truckenmiller A. The structure of oral language and reading and their relation to comprehension in kindergarten through grade 2. Read Writ. 2015;28(5):655-681. doi:10.1007/s11145-015-9544-5

29. Storch SA, Whitehurst GJ. Oral language and code-related precursors to reading: evidence from a longitudinal structural model. Dev Psychol. 2002;38(6):934-947. doi:10.1037/0012-1649.38.6.934

30. Nevo E, Vaknin-Nusbaum V. Enhancing language and print-concept skills by using interactive storybook reading in kindergarten. J Early Child Lit. 2018;18(4):545-569. doi:10.1177/1468798417694482

31. Isbell R, Sobol J, Lindauer L, Lowrance A. The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. Early Child Educ J. 2004;32(3):157-163. doi:10.1023/B:ECEJ.0000048967.94189.a3

32. Mol SE, Bus AG, de Jong MT. Interactive book reading in early education: a tool to stimulate print knowledge as well as oral language. Rev Educ Res. 2009;79(2):979-1007. doi:10.3102/0034654309332561

33. Krcmar M, Cingel DP. Parent–child joint reading in traditional and electronic formats. Media Psychol. 2014;17(3):262-281. doi:10.1080/15213269.2013.840243

34. Takacs ZK, Swart EK, Bus AG. Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: a meta-Analysis. Rev Educ Res. 2015;85(4):698-739. doi:10.3102/0034654314566989

35. Verhallen MJAJ, Bus AG. Video storybook reading as a remedy for vocabulary deficits: outcomes and processes. J Educ Res Online. 2009;1(1):172-196. doi:10.25656/01:4561

36. Shute VJ, Ventura M, Bauer MI, Zapata-Rivera D. Melding the power of serious games and embedded assessment to monitor and foster learning: flow and grow. In: Ritterfeld U, Cody MJ, Vorderer P, editors. Serious games: Mechanisms and effects. New York: Routledge; 2009:295-321. ISBN:0415993709

37. Nagle A, Riener R, Wolf P. Personality-based reward contingency selection: A player-centered approach to gameplay customization in a serious game for cognitive training. Entertain Comput. 2018;28:70-77. doi:10.1016/j.entcom.2017.10.005

38. van de Ven M, de Leeuw L, van Weerdenburg M, Steenbeek-Planting EG. Early reading intervention by means of a multicomponent reading game. J Comput Assist Learn. 2017;33(4):320-333. doi:10.1111/jcal.12181

39. Hsi S. Conceptualizing learning from the everyday activities of digital kids. Int J Sci Educ. 2007;29(12):1509-1529. doi:10.1080/09500690701494076

40. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. Front Psychiatry. 2017;7:209. doi:10.3389/fpsyt.2016.00209

41. McTigue EM, Solheim OJ, Zimmer WK, Uppstad PH. Critically reviewing GraphoGame across the world: recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. Read Res Q. 2019;55(1):45-73. doi:10.1002/rrq.256

42. Verhoeven L, Voeten M, van Setten E, Segers E. Computer-supported early literacy intervention effects in preschool and kindergarten: a meta-analysis. Educ Res Rev. 2020;30:100325. doi:10.1016/j.edurev.2020.100325

43. Scanlon M, Buckingham D, Burn A. Motivating maths? Digital games and mathematical learning. Technol Pedagog Educ. 2005;14(1):127-139. doi:10.1080/14759390500200187

44. Ameryoun A, Sanaeinasab H, Saffari M, Koenig HG. Impact of game-based health promotion programs on Body Mass Index in overweight/obese children and adolescents: a systematic review and meta-analysis of randomized controlled trials. Child Obes. 2018;14(2):67-80. doi:10.1089/chi.2017.0250

45. Zou D, Huang Y, Xie H. Digital game-based vocabulary learning: where are we and where are we going? Comput Assist Lang Learn. 2021;34(5-6):751-777. doi:10.1080/09588221.2019.1640745

46. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. New York: Harper Perennial; 1990. ISBN:0060162538

47. Geurts L, Wouters J. Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am. 1999;105(4):2476-2484. doi:10.1121/1.426851

48. Van Hirtum T, Moncada-Torres A, Ghesquière P, Wouters J. Speech envelope enhancement instantaneously effaces atypical speech perception in dyslexia. Ear Hear. 2019;40(5):1242-1252. doi:10.1097/AUD.0000000000000706

49. Van Herck S, Vanden Bempt F, Economou M, et al. Ahead of maturation: enhanced speech envelope training boosts rise time discrimination in pre-readers at cognitive risk for dyslexia. Dev Sci. 2021;00:e13186. doi:10.111/desc.13186

50. Koning R, Wouters J. The potential of onset enhancement for increased speech intelligibility in auditory prostheses. J Acoust Soc Am. 2012;132(4):2569-2581. doi:10.1121/1.4748965

51. Koning R, Wouters J. Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users. Hear Res. 2016;342:13-22. doi:10.1016/j.heares.2016.09.002

52. Goswami U. A temporal sampling framework for developmental dyslexia. Trends Cogn Sci. 2011;15(1):3-10. doi:10.1016/j.tics.2010.10.001

53. Goswami U, Thomson J, Richardson U, et al. Amplitude envelope onsets and developmental dyslexia: A new hypothesis. Proc Natl Acad Sci U S A. 2002;99(16):10911-10916. doi:10.1073/pnas.122368599

54. Glatz T, Vanderauwera, J, Vanden Bempt, F, et al. Design of GraphoGame Flemish. OSF; April 7, 2021. doi:10.17605/OSF.IO/CYKED

55. Vanden Bempt F, Van Herck S, Economou M, et al. Digital game-based phonics instruction promotes print knowledge in pre-readers at cognitive risk for dyslexia. Front Psychol. 2021;12:720548. doi:10.3389/fpsyg.2021.720548

56. Raven JC, Court JH, Raven J. Manual for Raven's Progressive Matrices and Vocabulary Scales. London: Lewis; 1984. ISBN:07186047179780718604714

57. Boets B, De Smedt B, Cleuren L, Vandewalle E, Wouters J, Ghesquière P. Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. Br J Dev Psychol. 2010;28(1):5-31. doi:10.1348/026151010X485223

58. Geurts L, Vanden Abeele V, Celis V, et al. Diesel-X: a game-based tool for early risk detection of dyslexia in preschoolers. In: Describing and studying domain-specific

serious games. Cham: Springer International Publishing; 2015:93-114. ISBN:9783319202754

59. van den bos KP, Zijlstra BJH, Spelberg HC. Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. Scienitific Stud Read. 2002;6(1):25-49. doi:10.1207/S1532799XSSR0601_02

60. Verwimp C, Vanden Bempt F, Kellens S, et al. Pre-literacy heterogeneity in Dutch-speaking kindergartners: Latent profile analysis. Ann Dyslexia. 2020;70(3):275-294. doi:10.1007/s11881-020-00207-9

61. Van Wieringen A, Wouters J. LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. Int J Audiol. 2008;47(6):348-355. doi:10.1080/14992020801895144

62. Unity. Unity Documentation - 2D or 3D projects. 2017. Available from: https://docs.unity3d.com [accessed Nov 13, 2021]

63. Unity Technologies. Unity Asset Store. 2021. Available from: https://assetstore.unity.com/ [accessed Nov 13, 2021]

64. Laranjeiro D. Development of Game-Based M-Learning Apps for Preschoolers. Educ Sci. 2021;11(5):229. doi:10.3390/educsci11050229

65. Clements DH, Sarama J. Teaching with computers in early childhood education: Strategies and professional development. J Early Child Teach Educ. 2002;23(3):215-226. doi:10.1080/1090102020230305

66. Kort W, Schittekatte M, Compaan E. CELF-4-NL: Clinical Evaluation of Language Fundamentals Nederlandse Versie. 4th edition. Amsterdam: Pearson; 2008.

67. Dunn LM, Dunn LM. Peabody Picture Vocabulary Test-III-NL PPVT-III-NL. Amsterdam: Harcourt; 2005.

68. Wiig EH, Secord WA, Semel E. CELF Preschool-2-NL: Clinical Evaluation of Language Fundamentals Preschool: Nederlandstalige Versie. 2nd edition. Amsterdam: Pearson; 2012. ISBN:9783319912790

69. Read JC. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. Cogn Technol Work. 2008;10(2):119-128. doi:10.1007/s10111-007-0069-9

70. R Studio Team. RStudio: Integrated development for R. 2019. Available from: http://www.rstudio.com/ [accessed Dec 15, 2021]

71. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2021. Available from: https://www.r-project.org/ [accessed Dec 15, 2021]

72. Komsta L. mblm: Median-based linear models. 2019. Available from: https://cran.r-project.org/web/packages/mblm/mblm.pdf [accessed Jun 15, 2021]

73. Sen PK. Estimates of the regression coefficient based on Kendall's Tau. J Am Stat Assoc. 1968;63(324):1379-1389. doi:10.1080/01621459.1968.10480934

74. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. J Stat Softw. 2017;82(1):1-26. doi:10.18637/jss.v082.i13

75. Perinelli E, Gremigni P. Use of social desirability scales in clinical psychology: A systematic review. J Clin Psychol. 2016;72(6):534-551. doi:10.1002/jclp.22284

76. Zaman B, Vanden Abeele V, De Grooff D. Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. Int J Child-Computer Interact. 2013;1(2):61-70. doi:10.1016/j.ijcci.2012.12.001

77. Shields BJ, Palermo TM, Powers JD, Grewe SD, Smith GA. Predictors of a child's ability to use a visual analogue scale. Child Care Health Dev. 2003;29(4):281-290. doi:10.1046/j.1365-2214.2003.00343.x

78. van Otterloo SG, van der Leij A, Veldkamp E. Treatment integrity in a home-based pre-reading intervention programme. Dyslexia. 2006;12(3):155-176. doi:10.1002/dys.311

79. Gersten R, Fuchs LS, Compton D, Coyne M, Greenwood C, Innocenti MS. Quality indicators for group experimental and quasi-experimental research in special education. Except Child. 2005;71(2):149-164. doi:10.1177/001440290507100202

80. Justice LM, Skibbe LE, McGinty AS, Piasta SB, Petrill S. Feasibility, efficacy, and social validity of home-based storybook reading intervention for children with language impairment. J Speech, Lang Hear Res. 2011;54(2):523-538. doi:10.1044/1092-4388(2010/09-0151)

81. Al Otaiba S, Fuchs D. Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. J Learn Disabil. 2006;39(5):414-431. doi:10.1177/00222194060390050401

82. Katzir T, Goldberg A, Aryeh TJB, Donnelley K, Wolf M. Intensity vs. duration: Comparing the effects of a fluency-based reading intervention program, in after-school vs. summer school settings. J Educ Train Stud. 2013;1(2):61-73. doi:10.11114/jets.v1i2.143

83. Boot WR, Simons DJ, Stothart C, Stutts C. The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. Perspect Psychol Sci. 2013;8(4):445-454. doi:10.1177/1745691613491271

84. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998;2(14):1-74. doi:10.3310/hta2140

85. Van Kleeck A, Vander Woude J, Hammett L. Fostering literal and inferential language skills in head start preschoolers with language impairment using scripted book-sharing discussions. Am J Speech-Language Pathol. 2006;15(1):85-95. doi:10.1044/1058-0360(2006/009)

86. Tompkins V, Bengochea A, Nicol S, Justice LM. Maternal inferential input and children's language skills. Read Res Q. 2017;52(4):397-416. doi:10.1002/rrq.176

87. Pena ED. Measurement of modifiability in children from culturally and linguistically diverse backgrounds. Commun Disord Q. 2000;21(2):87-97. doi:10.1177/152574010002100203

Abbreviations

AC	Active control
AC_NE	Active control non-enhanced group
ADHD	Attention Deficit Hyperactivity Disorder
ATH	Audiotechnica
DLD	Developmental language disorder
DYSCO	Dyslexia Research Collaboration
EE	Envelope enhancement
GG-FL	GraphoGame Flemish
GG-FL_EE	GraphoGame Flemish envelope enhanced group
GG-FL_NE	GraphoGame Flemish non-enhanced group
KU Leuven	Katholieke Universiteit Leuven (Catholic University Leuven)
QRA	Question response accuracy
RCT	Randomized controlled trial
SES	Socio-economic status
TD	Training duration
UZ Leuven	Universitair Ziekenhuis Leuven (University Hospital Leuven)

Multimedia Appendix 1: Technical story game development guide

Short game description and rationale

The story game is a tablet application, originally developed to test the efficacy of listening to envelope enhanced speech (EE) on phonology and reading via improved basic auditory speech processing among five-year old children at cognitive risk for dyslexia. The EE algorithm automatically detects and amplifies important rhythmic acoustic cues of the speech envelope in digital audio recordings, i.e., onset rise times [1–5], which are considered to be important for phonological skill development [6,7]. The story game mainly consists of a main intervention task, i.e., listening to Dutch-spoken recorded age-appropriate children stories, followed by a story rating part and a short quiz with basic content-related questions. In total, the game provides 72 story sessions, categorized in 18 game phases of four grouped sessions containing four (longer) to eight (shorter) stories from the same book series and author. The game facilitates rewards in the form of stars merely for the full completion of a game session and in the form of coins for responding correctly to the content-related questions. Players can spend their earned coins in a virtual shop, i.e., the avatar customization system. The shop items customizes the players' avatars, who move around a map, i.e., the virtual hub world, and inform players on their game progress. The game is suitable for Android Devices (in the case of the current study: Samsung Galaxy Tabs E9.6). Since the game is intended for five-year-old children, it has been developed as functionally simple as possible. For instance, it does not offer any form of written text and the players are only required to listen to spoken text and press buttons which appear vividly on the screen.

Game development

Pre-programming phase

Prior to the programming phase, a story game concept, which consisted of game-based story listening phases, was created by different members of the research group, taking into account the elements of the GameFlow model (except for the 'social interaction' element) [8]. The reason for not including the social interaction element was driven by the fact that the participants of the current study were too young for between-player interactions (e.g., via chat boxes) and that all participants were anonymously and independently recruited, hindering communication between different players. After finalizing the concept, several children storybooks were screened via library and publisher visits. Storybooks were eligible for the story game (1) if they were part of a book series that contained minimally three other published books, (2) when their target age was four to six years old, and (3) if they lasted approximately five to fifteen minutes. The limitation in length was set based on similar story lengths in previous kindergarten story listening studies [9,10] and because of the feasibility for young children to combine the story game with the other intervention games for 12 weeks on a nearly daily basis. Based on these criteria, we selected 87 storybooks from 14 different book-series, for which all publishers sent written consents via email to use the book texts and images within the framework of the intervention study (see Supplementary Table 1 for an overview of the game phases and their corresponding book series, authors, illustrators, and publishers). For each story session, the research group created three content-related multiple-choice questions with three response alternatives each. All selected stories were then recorded by nine different highly skilled female speakers with knowledge of Standard Dutch in Flanders. Female voices were opted for since a previous study showed better learning outcomes in female-instructed digital problem-solving tasks and therefore advised the use of female voices when designing multimedia learning environments [11]. Yet, to bring some variation into the speakers' voices during the game sessions, the game instructions and content-related questions and their corresponding multiple-choice alternatives were recorded

by a male Flemish radio presenter and a female eight-year old child respectively. Next, all recordings were edited to exclude errors and slip of the tongues. The envelope enhancement on the story recordings for the GG-FL_EE group was done by a signal processing expert and is fully described in the study of Van Herck and colleagues [2]. For each story, the corresponding story illustrations from the book were linked to a specific timeframe of the audio recording and summarized in so-called audiovisual-synchronization documents. Incorporating images along with the story recordings was based on the idea that illustrations benefit narrative comprehension [9,12] and foster engagement [13]. The story recordings, illustrations and their accompanying question recordings and audiovisual-synchronization information were ordered in a predetermined file structure (see rationale and clarification of this file structure in the section on the programming details of the main intervention task environment).

Supplementary Table 1. Book series, authors, illustrators, and publishers of the stories implemented in the story game.

Game phase	Session	Book series	Author(s)	Illustrator	Publisher	Amount of stories
1	1 – 4	Lotta de Kip	Diane Put, Rik De Wulf	Rik De Wulf	Clavis	8
2	5 - 8	Ella het heksje	Guy Daniëls	Lisa Brandenburg	Clavis	4
3	9 – 12	Avonturen van Tommie en Lotje	Jacques Vriens	Kees de Boer	Van Holkema& Warendorf	4
4	13 – 16	Joep en Pleuntje 1	Isabelle De Ridder	Monique Dozy	Clavis	4
5	17 – 20	Winnie de Poeh	A.A. Milne	Ernest H. Shepard	Deltas	4
6	21 – 24	Het Winkelstraatje	Marianne Busser, Ron Schröder	Ingrid ter Koele	Moon	4
7	25 – 28	Verhalen uit het Elzen-Eikenbos	Fiona Rempt, Noelle Smit	Noelle smit	Van Holkema& Warendorf	4
8	29 – 32	Beroepenserie 1	Liesbet Slegers	Liesbet Slegers	Clavis	6

9	33 – 36	Rikki	Guido Van Genechten	Guido Van Genechten	Clavis	8
10	37 – 40	Grote Anna 1	Kathleen Amant	Kathleen Amant	Clavis	4
11	41 – 44	Meneer Big	Gitte Spee	Gitte Spee	Pimento	4
12	45 – 48	Joep en Pleuntje 2	Isabelle De Ridder	Monique Dozy	Clavis	4
13	49 – 52	Beroepenserie 2	Liesbet Slegers	Liesbet Slegers	Clavis	7
14	53 – 56	Grote Anna 2	Kathleen Amant	Kathleen Amant	Clavis	4
15	57 – 60	Hekselien	An Melis	An Melis	Abimo	4
16	61 – 64	Sjon en Sjaan	Gitte Spee	Gitte Spee	Pimento	4
17	65 – 68	Beroepenserie 3	Liesbet Slegers	Liesbet Slegers	Clavis	6
18	69 - 72	Lars de kleine ijsbeer	Hans De Beer	Hans De Beer	Vier Windstreken	4

Scripts

The story game was entirely programmed in Unity 3D [14] using the C# scripting language. Most of the inner workings of the game were detailed in appropriately named scripts. The scripts could generally be divided into four categories (see Supplementary Table 2 for an overview of all scripts implemented in the game):

- **Component scripts** handled specific situations, such as the shop in the avatar customization system, the avatar, the quiz, the ending animation, etc. They were generally used in one specific situation in the game.
- **Global scripts** persisted between different game phases (virtual hub world, main menu, shop in avatar customization system, and questions). Each script handled features that needed to be called in different situations spread throughout the game, i.e., saving, playing music, and playing instruction sounds.
- **Utility scripts** provided functions that were not specific to the game, but that were generally useful. For example, it included a script for orbiting the camera with touch.

- **Editor-specific scripts** existed to facilitate the game development phase and were not included in the game builds. They consisted of helper codes that did not run on the tablets of the participants, but solely on the computer of the game developer.

Supplementary Table 2. Scripts implemented in the story game

Script		Description
Global scripts		
	Global	This was an overall manager script, providing connections to all existing global scripts, facilitating loading, playing music, etc.
	Calibrator	Handled sound calibration user interface and applied those settings to the Unity Audio Mixer system.
	CharacterGenerator	Used saved settings to generate an instance of the customizable players' avatars.
	Configurator	Handled game configuration, i.e., saving, loading profiles, and changing profile variables such as the maximum number of sessions per 24 hours.
	Instructor	Played instruction audio, provided helper functions to select appropriate instruction audio.
	LevelData	Determined the colors of stars and stages in the different game phases.
	Music	Played and sustained music clips depending on the game context. Allowed sound to transition smoothly.
	SaveData	Saved and loaded profile specific data, such as stars, coins, unlocked levels, and equipment.
Components scripts		
	Camera Movement	Handled camera movement in the virtual hub world.
	Character	Controlled single avatars: their equipment, animations, and movements around the virtual hub world.
	EndingScene	Used to play the ending animation in the hub world, when a player finished the entire story game.
	MainMenu	Controlled user interface of the main menu.
	Map	Directed the game during the virtual hub world mode. Received player input and triggered transitions and animations.
	Quizmaster	Directed the game during the quiz mode. Guided question audio, buttons, animations, and transitions.

	StoryTeller	Directed the game during story listening mode. Guided story audio, panels, pausing, and triggered the quiz when the story was finished.
	Shop	Directed the game during the avatar customization mode. Guided buying of items, sound effects, music playing, and avatar swapping.
	ShopDisplayItem	Controlled props in the avatar customization system background, which appeared when the player was able to buy new items from the shop.
	ShopItem	Controlled a single user interface in the shop that could be unlocked, bought, and equipped. It included characters and accessories.
	StoryHolder	Held a set of stories, including the associated audio, questions, image panels, timings.
Utility scripts		
	CameraOrbit	Orbited the camera around a fixed point based on the touch input. This only occurred in the avatar customization system.
Editor-specific scripts		
	objExporter	Exported a Unity 3D model to a standardized 3D model format (.obj), which facilitated editing the models with other software.
	AutoLevel	Allowed the developer to directly end up in the avatar customization system, question part, or virtual hub world, without having to pass through the main menu.
	Recenter	Used to automatically reset positions of many objects, i.e., the animation models in the virtual hub world, facilitating positioning of different animations.
	StoryLoader	Loaded story data from the Stories folder and exported the data to Storyholder objects for easy access.

Specific design and programming details

Main intervention task environment

Design details

In the main intervention task environment, players were required to listen to one or two short stories, followed by a story rating and content-related questions. The story and accompanying imagery played on a mostly neutral black screen to lay focus on the content of the story itself. The story was paused at key moments, while displaying a large green play

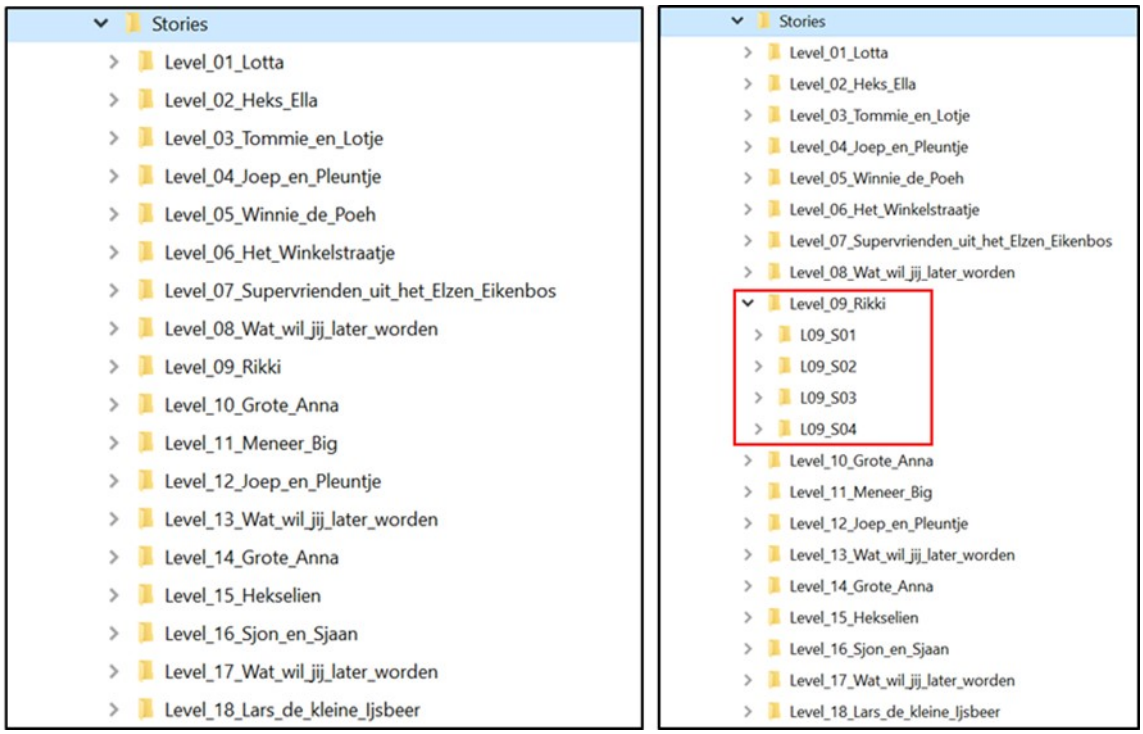
button to continue. This allowed a short break during the listening phase and ascertained that the player continued to pay attention to the story, since there was no option to rewind. While the recording played, images smoothly faded in and out. Once a single story finished, the story rating part and quiz occurred. Each time the question audio of the quiz was played, buttons representing the response options appeared on the screen. The response audio was played one by one, while highlighting each individual corresponding button. Once the response audio was finished, an extra replay button appeared to play the question and the response audio again, using the same animations. The buttons became intractable once the first round of audio was complete. The players immediately received feedback whether or not their answer was correct. A correct answer resulted in a short animation showing clearly that the player had earned a new coin. A wrong answer resulted in a short animation which revealed the correct response button. The response buttons had a simple design, based on the star theme of the game and showed either one, two, or three dots, representing response option one, two, or three respectively.

Programming details

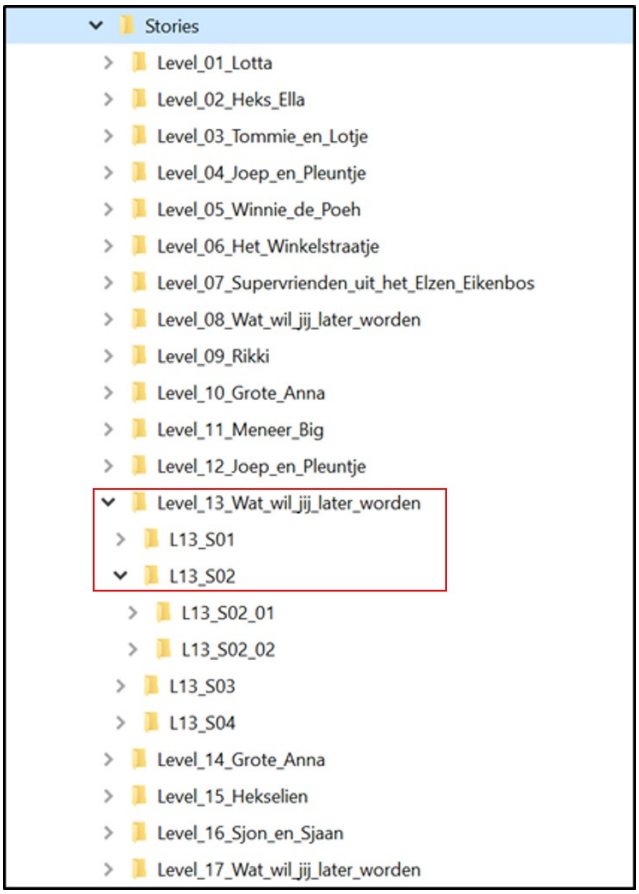
Story implementation

The stories in the game were subject to a fairly complex data structure, due to limitations of memory on Android. The story audio and accompanying images, the audiovisual-synchronization document, the question and response audio were ordered in a predetermined file structure and implemented in the game based on a specific format. This predefined file structures could automatically be loaded into the Unity editor based on the "*StoryLoader*" script. Thus, it accelerated the design process as it allowed to quickly implement updates regarding audiovisual synchronization documents, pauses, and questions, without extra specialized programming work. The main folder was named "*Stories*" (left

panel of Supplementary Figure 1) and contained 18 sub-folders, representing the different story themes of the 18 game phases (for example, the folder of game phase 9 with stories of Rikki the Rabbit was named "*Level_09_Rikki*"). Each game phase folder was further subdivided in four sub-phase folders (e.g., "*L09_S01*", "*L09_S02*", "*L09_S03*", "*L09_S04*", each containing a set of stories (either one long story or two short ones) the player should listen to in one game session (right panel of Supplementary Figure 1). In total, 72 sub-phase folders were created, e.g., four for each game phase, representing the total amount of 72 game sessions. When a sub-phase contained two short stories instead of one long, the folder was further subdivided into separate story folders, such that each story had its own folder (e.g., "*L13_S02_01"* and "*L13_S02_02"*) (Supplementary Figure 2). In total, 87 story folders were created, since a total of 87 stories were implemented in the game.



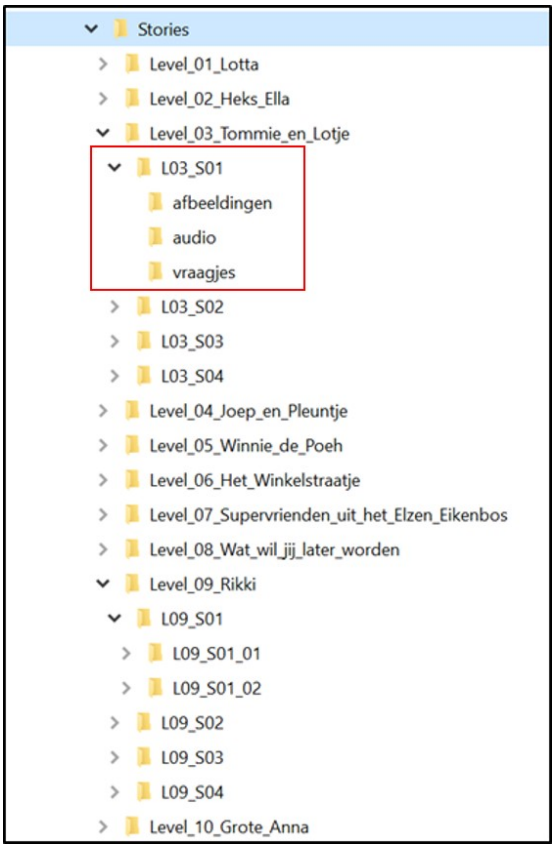
Supplementary Figure 1. File structure of the stories of the story game. Left panel) Overall game phase folders. Right panel) Sub-phase folders



Supplementary Figure 2. Example of a sub-phase folder (*L13_S02*) containing two separate story folders (*L13_S02_01 and L13_S02_02*).

Each of the 87 separate story folders contained three types of subfolders (Supplementary Figure 3):

- "*audio*" folder, containing the story recording in .wav format.
- "*afbeeldingen*" folder, consisting of (1) the set of images in .jpg format to be displayed during the story and (2) an audiovisual-synchronization .txt document (see detailed audiovisual synchronization section below). Each image was named numerically as "*afbxx.jpg"*. The .jpg format was preferred since it saved memory.
- *"vraagjes"* folder, containing the question and response recordings in .wav format along with a .txt document which specified the order of the questions and corresponding response alternatives (see detailed questions section below).



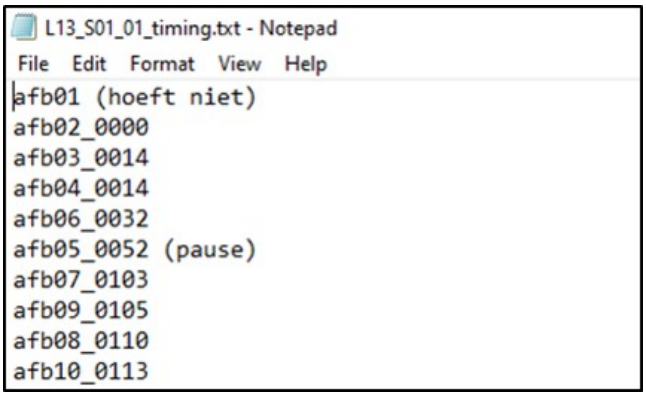
Supplementary Figure 3. The data structure of a story folder

The editor-specific "StoryLoader" script provided functions for importing data in the main "*Stories*" folder and sorting them into prefabs that the game used to load the story data. This setup was necessary, since otherwise the programmer was forced to load all of the story data at once to play a single story. The parsing assumed that the naming conventions outlined previously were being upheld. As such, everything could be imported, but also easily updated if necessary without issues.

Audiovisual synchronization

Each "*afbeeldingen*" folder of a story contained a so-called audiovisual-synchronization file in .txt format (see Supplementary Figure 4 for an example). It defined when image panels accompanying the story audio had to be shown. The file contained a list of the times at which the panel should show a particular image, specified as "*afbxx_mmss*". The

time was thus listed as a minutes-seconds format. For instance, the line "*afb09_0105*" in Supplementary Figure 4 meant that the image 'afb09.jpg' was shown at minute 1 and 5 seconds, or 65 seconds throughout the story. Timing regulations were programmed as follows: one second before a specific timing indication, the previous image panel faded out, at the timing per se, a black screen was shown, and one second after the timing, the image accompanying the specific timing appeared on the screen. When a timing indication was followed by the text "(*hoeft niet*)", the given image was ignored. A "(*pause)*" specification indicated that the given transition should also pause the story, prompting the player to continue it manually by pressing on the green arrow. If a given list of timings did not have any*"(pause)"* flags, the story paused after every single image by default.



Supplementary Figure 4. Example of an audiovisual synchronization document.

Questions

Each "*vraagjes*" folder contained a document which specified the order of the questions for the particular story and their corresponding response alternatives. It was consistently named as "*Storyname_vragen.txt*" (e.g., *L03_S01_vragen.txt* for the questions of the story in the first sub-phase of the third game phase) (Supplementary Figure 5). This file first specified the name of the question audio (e.g., L03_S01_vraag01 is the first question for the story), followed by a written example of the actual question and three possible answers, rounded off by a white line. Specifying the text content of the question and responses in these files was not really relevant to the game, but it was used for data exportation on the accuracy

of the questions. For each question, the correct answer was marked using the keyword *"(juist)".* The relevant audio corresponded the question name as specified in the question files. For instance, the question *"L03_S01_vraag01"* had an audio file which was named *"L03_S01_vraag01.wav"*.


L03_S01_vragen.txt - Notepad

File Edit Format View Help

```
L03_S01_vraag01
Waarom moet Lotje in bed blijven?
Ze is stout geweest
Ze is ziek (juist)
Het is midden in de nacht

L03_S01_vraag02
Waarom is Lotje verdrietig?
Ze heeft honger
Ze heeft eng gedroomd
Ze mag niet in de sneeuw spelen van mama (juist)

L03_S01_vraag03
Wat zit er in de doos van Tommie?
Een hoopje sneeuw (juist)
Bloemen
Steentjes
```

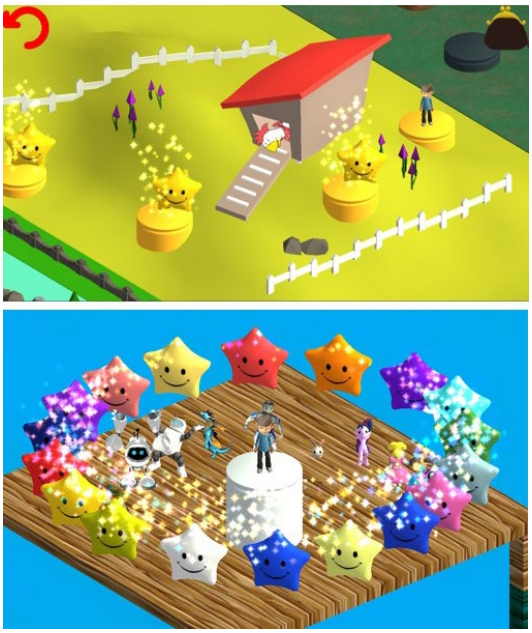
Supplementary Figure 5. Example of a question file. L03_S01_vragen.txt

Virtual hub world

Design details

The virtual hub world showed a visual overview of the current story theme (e.g., one of the 18 game phases) as well as the overall progress of the player. A customizable avatar jumped from one cylindrical stage to the next, when finishing a story session. Each stage thus represented one story session. The virtual hub world was shown in a restrictive orthogonal perspective, which revealed only a handful of stages at a time. The reasoning behind this was to create anticipation towards what appeared next on the map and to motivate players to come back to see what would be revealed next. The virtual hub world was designed in a way that the art work matched the content of the stories (e.g., a chicken coop representing the home of the main character in the four story sessions with books of "Lotta De Kip"/ "Lotta the

Chicken", as shown in the upper panel of Supplementary Figure 6). When a story became available, a big green flashing button appeared on the screen, inviting the player to listen to it. The hub world also contained the ending animation, which was shown when all game sessions were completed. It brought together all the elements of the game, including all possible game avatars and one star from each of the 18 game phases (see lower panel of Supplementary Figure 6).



Supplementary Figure 6. Upper panel) Alignment of the hub world with the themes of the story series. Lower panel) Ending cut-scene in the hub world.

Programming details

Supplementary Table 3 provides an overview of the art work (props) implemented in the virtual hub world and their corresponding sources.

Supplementary Table 3. Props and their sources implemented in the virtual hub world.

Props	Source
Forests	Unity Asset Store: Free low Poly Nature Pack [15]
Fall forest & woodland textures	Unity Asset Store: Low Poly Nature Pack Lite [16]
City buildings	Recycled from Diesel-X [17]
Spooky mansion	Sketchfab: World Skills Spooky Mansion by Philip Osborne [18]

Starting area	Unity Asset Store: Low Poly Style Environment [19]

Avatar customization system

Design details

After the main intervention task (e.g., listening to one or two stories, the story rating, and the short content-related quiz), players automatically ended up in a so-called avatar customization system, where they could spend earned coins on new avatars and/or accessories for their existing avatars in a little shop. The ten possible avatars (i.e., young boy/girl, robot, dragon, farm animal, pony, popstar, princess, knight, American Indian, or astronaut) were not directly linked to the story contents of the game, but they had been selected based on (1) their attraction both for boys and girls, (2) the ease for game implementation, and (3) the possibility to deck them out with plenty of accessory options. Each avatar could be unlocked based on a certain amount of earned coins and had eight distinct accessories that could be equipped simultaneously. Once players bought avatars or accessories, it was displayed on the screen. Two types of avatars were considered as special cases. First, as from the start of the intervention, the young boy or girl avatar (depending on the gender of the player for which an account was created) was set as the default initial avatar. Second, the farm animal used its accessory options to swap the current farm animal model for a different animal. It also had more options than any other avatar. Prices for all the avatars and accessories were set so that a perfect player could still spend all of his/her coins in the avatar customization shop.

Programming details

The code for implementing avatars was streamlined as much as possible, so that new avatars could be added to the game with only little extra programming effort. All avatars were 3D models (either custom-made, acquired from Sketchfab, or the Unity asset Store) with

textures (coloration) and rigging (tools which allowed the avatar to move in the game). For each avatar, the following properties were separately programmed:

- An animation system, which allowed different movements of the avatars, containing four possible animations:
    - Idle: looping animation that played by default when the avatar was in a resting state (e.g., when showed in the main menu)
    - Jump: played when a jump from one cylindrical stage to the next started. The animation played in reverse when the jump ended and the avatar landed.
    - Airborne: looping animation that played in the middle of a jump
    - StarPose: animation that played directly after a new star appeared on the cylindrical stage that was just abandoned.
- A shop entry for the avatar, which included its pricing and a 256x256 preview image
- A set of accessories connected to the avatar. For each accessory, two extra properties were programmed.
    - A shop entry for each unlockable accessory, which included its pricing and a 100x100 preview image
    - An optional 'ShopDisplayItem'. This was a non-functional version of the item that was visible in the shelves of the shop once it had been unlocked.

Supplementary Table 4 presents a list of the ten avatars, their accessories, and their sources. Supplementary Table 5 provides an overview of all sources of the available accessories. Accessories not mentioned in Supplementary Table 5 were custom made for the project.

Supplementary Table 4. Avatars and sources

Avatar	Accessories	Source

Boy/girl (default)	Propeller or sun hat, backpack, skates, special outfit, other hair color, party hat, swimming floaty, outfit change	Unity Asset Store [21]
Robot	Red paint, happy face, angry face, jetpack fire, mini robot, surprised face, shield bubble, blue paint	Unity Asset Store [22]
Astronaut	Green paint, robot arm, satellite dish, laser sword, star badge, screen on belly, rocket boots, star necklace	Custom made
Dragon	Red dragon, white dragon, propeller hat, winged shoes, mini dragon, dog collar, big wings, saddle	Unity Asset Store [23]
Knight	Sword, shield, helmet plume, crown, sparkles, red armor, green armor, star necklace	Unity Asset Store [24] (not available online anymore)
Farm animals	Variations in animal models act like accessories: duck, chicken, cat, dog, pig, sheep, goat, horse, lama, cow	Unity Asset Store [25]
Pony	Horn, open wings, sparkles, red pony, rainbow pony, star necklace, rainbow bracelets, mini pony	Sketchfab [26]
Princess	Skirt ribbon, rose, fairy hat, crown jewel, magic wand, blue outfit, star necklace, sparkles	Custom made
Popstar	Sunglasses, microphone, new guitar, flying speakers, rainbow spotlight, red outfit, black outfit, sparkles	Custom made
Native American	Feather, chieftain hat, bow, necklace, parrot, dream catcher, quiver, outfit change	Custom made

Supplementary Table 5. Accessories and sources

Accessory	Source
Backpack	Included in asset pack of boy/girl avatar
Knight crown	Sketchfab
Princess crown	Sketchfab
Dream catcher	Sketchfab
Swimming floaty	Sketchfab
Fairy hat	Sketchfab
Knight sword	Separated from base model
Knight shield	Separated from base model
Laser sword	Sketchfab
Mini robot	Sketchfab
Sun hat	Included in asset pack of boy/girl avatar
Quiver	Sketchfab
Rose	Sketchfab
Pony horn	Separated from base model
Pony open wings	Separated from base model

Additional game features

Main menu

The main menu of the story game appeared as a static scene of a bedroom and menu elements were designed into a book, to imitate a bedtime story moment (Supplementary Figure 7). The artwork of the fairytale book and bedroom elements were acquired from Sketchfab. The main menu showed the active player account and the amount of total earned stars to inform the player on the progress throughout the game. The settings system, represented by the gear wheel icon, was locked with a password to prevent regular users from accessing it. A game session began by pressing on the large green arrow. To exit the game, players were required to press the red arrow.



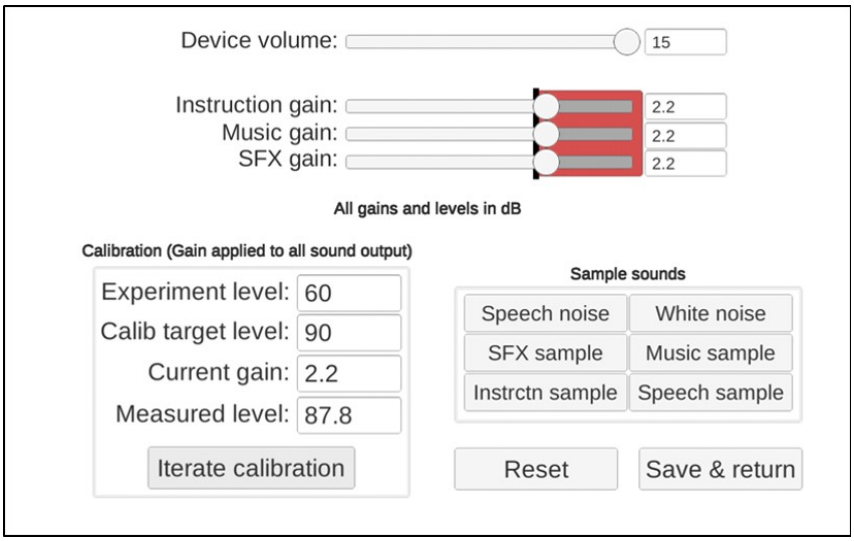
Supplementary Figure 7. Main menu of the story game.

Settings

Sound-related settings

Sound-related settings included options to calibrate the sound level of the stories in the game, and adjust the sound level of the sound effects, instructional voices, and music tracks (see Supplementary Figure 8). The 'Device volume' slider controlled the Android volume setting which could otherwise be adjusted via the buttons on the side of the tablet. This input was restricted during story game play to make sure players could not interfere with the

calibration. The other three sliders controlled gain values in dB for instruction speech (cfr. Instruction gain), music (cfr. Music gain) and sound effects (cfr. SFX gain) respectively. These were set by default to an agreed upon level, but could be adjusted if necessary. The calibration panel was used for calibrating the sound levels of the stories in the game. Story sound levels were calibrated to be played at 60 dB-A (see "*Experiment level*" box in Supplementary Figure 8). Calibration was performed in a sound proof booth by playing speech weighted noise that was constructed based on the average spectra of the nine different female story tellers at 90 dB-A (i.e., see "*Calib target level box*" in Supplementary Figure 8) through a headphone (ATH m20x). The measured output level was then filled in in the "*Measured level*" box and the resulting gain, specified in the "*Current gain"* box was automatically applied to the story recordings, but also to the instruction, music and sound effect gains. Calibration was iterated as long as necessary. The sample sounds panel allowed to play example recordings of the story speech sound (i.e., speech sample in Supplementary Figure 8), sound effects (SFX sample in Supplementary Figure 8), and music (Music sample in Supplementary Figure 8) to check whether the applied gains after calibration were acceptable. It also provided short audio recordings of white noise and the speech weighted noise, which was used for calibration of the story sound levels.



Supplementary Figure 8. Sound-related settings

Game configuration settings

The game configuration settings (Supplementary Figure 9) offered a series of settings that could be used to adjust a certain game account behavior (consider Supplementary Table 6 for an overview of the different game settings). Game configuration settings also included panels to add or remove player accounts. Each account was recognized based on a name and ID number (e.g., Name-01 as specified in Supplementary Figure 9). The name was displayed to the player in the main menu, while the ID number was used internally by members of the research group to recognize and differentiate large numbers of test subjects. Experimental results of each account (e.g., play dates and hours, game progress, question response accuracy, and story ratings) were saved as binary .sav files. On Android, such files were stored in the data folder, under the package name for the game, which was set as "*be.kuleuven.med.exporl.InterventionGame*" in the current intervention study. Manually exporting the data (as specified by the "*Export now*" button in Supplementary Figure 9) made sure that the .sav files were stored in an easily accessible folder on the Android device, which was created by members of the research group and for which its location was specified by the path in the configuration panel. Experimental play data were also converted in a user-readable format, i.e., a text file (.txt) for each individual player containing detailed player information, and a simplified csv file (.csv) containing more general play data.



Supplementary Figure 9. Game configuration settings

Supplementary Table 6. Available game configuration settings

Setting name	Purpose
Restrict number of plays	Blocked player from playing more than one game session per 24 hours.
Stockpile missed plays	Allowed players to catch up if they forgot to play a session, e.g., they could play three story sessions in a row if they forgot two days of playing.
Stories per day	The amount of story sessions that could be played per day. Note that "1" was set as default in the intervention study.
Starting coins	The amount of coins that were initially given to a newly created profile. This setting did not affect existing profiles. It was intended for debugging purposes and was set as "0" in the current intervention study.
Starting stars	The amount of stars that were initially given to a newly created profile. This setting did not affect existing profiles. It was intended to let players start at different game sessions in the game, but it was also useful for debugging purposes. In the current intervention study, it was always set as "0" so that every player started with the first story of the first sub-phase of the first game phase.
Reset hour	The hour at which the restriction counting was triggered. For instance, if this value was set as 04:00:00, the player was blocked from playing a second game session until 4 AM.

Show debug tools	Displayed tools for fast forwarding and skipping trough stories. It was useful for faster debugging, but was turned off in the current intervention study.

Music and sound effect implementation

Music tracks implemented in the game were royalty free music written by Kevin Macleod [27] (see Supplementary Table 7 for the track titles and the occasion when they occurred in the game). Sounds effects where acquired from a freely accessible website and edited where necessary [28]. Supplementary Table 8 provides an overview of the different sound effects and their occurrences in the game.

Supplementary Table 7. Track names and occurrence in the story game.

Track name	Occurrence in game
Melodie Victoria	Main menu
Carefree	Virtual hub world
Cheery Monday	Virtual hub world
Teddy Bear Waltz	Virtual hub world
Zazie	Avatar customization system

Supplementary Table 8. Sound effects and their occurrence in the story game.

Sound effect	Occurrence in the game
Button	Occurs when a player presses on the green arrow to start the game session in the main menu
Buy	Occurs when a player buys a new avatar or accessory
Chime	Occurs when a new story starts
Coin	Occurs when opening the avatar customization system. It represents the sound of falling coins
Correct	Occurs when the answer to a content-related question is correct
Equip	Occurs when an accessory is applied to an avatar
Error	Occurs when a player tries to buy accessories/avatars without enough coins
Select	Occurs when avatars change
Unequip	Occurs when an accessory is removed from the avatar
Wrong	Occurs when the answer to a content-related question is incorrect

Exporting data experimental results to server

Experimental results (e.g., the generated .txt and .csv files for each account) were automatically logged to a University internet server, so that data were backed up during the home-based intervention and members of the research group could keep track of the intervention trajectory of all participants for a longer period of time. Collecting and synchronizing these experimental results was handled by an external so-called Android AutoSyncer application, which was specifically developed for the current intervention study and installed on all tablets. The AutoSyncer application could communicate with a basic server over the internet to send data files and sort them into a predefined data structure, thus maintaining in-game player data from many different users at the same time.

Bugs

Failure to delete active profiles

When a player exited the game via the intended red arrow in the main menu, the game saved the currently active profile one last time. Yet, if that active profile had been previously deleted for some reason in the main configuration system, it would just be rewritten back, making it seemingly impossible to delete that profile. This bug could be circumvented by closing the game externally, for example via the Android task manager, yet it should be fixed in future versions of the story game.

Blank name saving and related “Bobby bug”

Account-related .sav files usually contained a name structure, such as “*Name-ID.sav*”. However, they were sometimes inexplicably saved without any name or ID specification, e.g., “.sav” (i.e., blank name saving). This bug messed with any code trying to save, load or delete the existing account-related *Name-ID.sav* files and sometimes even replaced the existing *Name-ID.sav* file on the tablet, causing loss of data. Moreover, this blank name saving bug

sometimes resulted in the occurrence of another bug, i.e., the so-called “Bobby bug”. More specifically, when the code failed to load an account-related profile, the name “Bobby” appeared on the main menu screen, instead of the name of the participant’s account, as it was used as a place holder name to test the user interface of the game in the developmental phase. Moreover, in this case, the story game only played the first sub-phase of the first game phase over and over again. Occurrence of these bugs in the current intervention study resulted in a number of home visits where members of the research group fully reset the entire game and existing profiles. Yet, thanks to the existence of the AutoSyncer application, this bug only caused a limited amount of data loss (i.e., in seven participants). However, when applying the game in future studies, fixing these bugs must take the highest priority.

References

1. Geurts L, Wouters J. Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am. 1999;105(4):2476-2484. doi:10.1121/1.426851

2. Van Herck S, Vanden Bempt F, Economou M, et al. Ahead of maturation: enhanced speech envelope training boosts rise time discrimination in pre-readers at cognitive risk for dyslexia. Dev Sci. 2021;00:e13186. doi:10.111/desc.13186

3. Van Hirtum T, Moncada-Torres A, Ghesquière P, Wouters J. Speech envelope enhancement instantaneously effaces atypical speech perception in dyslexia. Ear Hear. 2019;40(5):1242-1252. doi:10.1097/AUD.0000000000000706

4. Koning R, Wouters J. Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users. Hear Res. 2016;342:13-22. doi:10.1016/j.heares.2016.09.002

5. Koning R, Wouters J. The potential of onset enhancement for increased speech intelligibility in auditory prostheses. J Acoust Soc Am. 2012;132(4):2569-2581. doi:10.1121/1.4748965

6. Goswami U. A temporal sampling framework for developmental dyslexia. Trends Cogn Sci. 2011;15(1):3-10. doi:10.1016/j.tics.2010.10.001

7. Goswami U, Thomson J, Richardson U, et al. Amplitude envelope onsets and developmental dyslexia: A new hypothesis. Proc Natl Acad Sci U S A. 2002;99(16):10911-10916. doi:10.1073/pnas.122368599

8. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. ACM Comput Entertain. 2005;3(3):1-24. doi:10.1145/1077246.1077253

9. De Jong MT, Bus AG. Quality of book-reading matters for emergent readers: an experiment with the same book in a regular or electronic format. J Educ Psychol. 2002;94(1):145-155. doi: 10.1037/0022-0663.94.1.145

10. Piotrowski JT, Krcmar M. Reading with hotspots: young children's responses to touchscreen stories. Comput Human Behav. 2017;70:328-334. doi:10.1016/j.chb.2017.01.010

11. Linek SB, Gerjets P, Scheiter K. The speaker/gender effect: does the speaker's gender matter when presenting auditory text in multimedia messages? Instr Sci. 2010;38(5):503-521. doi:10.1007/s11251-009-9115-8

12. Mayer RE. Cognitive Theory of Multimedia Learning. In: Mayer RE, editors. The Cambridge Handbook of Multimedia Learning. Cambridge: Cambridge University Press; 2005:31-48. ISBN:9781107035201

13. Laranjeiro D. Development of Game-Based M-Learning Apps for Preschoolers. Educ Sci. 2021;11(5):229. doi:10.3390/educsci11050229

14. Unity. Unity Documentation - 2D or 3D projects. 2017. Available from: https://docs.unity3d.com [accessed Nov 13, 2021]

15. Unity Technologies. Unity Asset Store: Low-Poly Simple Nature Pack. 2021. Available from: https://assetstore.unity.com/packages/3d/environments/landscapes/low-poly-simple-nature-pack-162153 [accessed Nov 12, 2021]

16. Unity Technologies. Unity Asset Store: Low Poly Nature pack (Lite). 2021. Available from: https://assetstore.unity.com/packages/3d/environments/landscapes/low-poly-nature-pack-lite-40444 [accessed Nov 12, 2021]

17. Geurts L, Vanden Abeele V, Celis V, et al. Diesel-X: a game-based tool for early risk detection of dyslexia in preschoolers. In: Describing and studying domain-specific serious games. Cham: Springer International Publishing; 2015:93-114. ISBN:9783319202754

18. Sketchfab. Spooky Mansion 3D Model. 2021. Available from: https://sketchfab.com/3d-models/spooky-mansion-ccb25ffdcf3047f9a9c136faab8e3e00 [accessed Nov 12, 2021]

19. Unity Technologies. Unity Asset Store: Low Poly Style Environment. 2021. Available from: https://assetstore.unity.com/packages/3d/environments/low-poly-style-environment-72471 [accessed Nov 12, 2021]

20. Unity Technologies. Unity Asset Store: Low Poly Style Environment. 2021. Available from: https://assetstore.unity.com/packages/3d/environments/low-poly-style-environment-72471 [accessed Nov 12, 2021]

21. Unity Technologies. Unity Asset Store: 2 toon kids. 2021. Available from: https://assetstore.unity.com/packages/3d/characters/humanoids/2-toon-kids-104298 [accessed Nov 12, 2021]

22. Unity Technologies. Unity Asset Store: Robot B-127. 2021. Available from: https://assetstore.unity.com/packages/3d/characters/robots/robot-b-127-72406 [accessed Nov 12, 2021]

23. Unity Technologies. Unity Asset Store: Little Dragons: Sea. 2021. Available from: https://assetstore.unity.com/packages/3d/characters/creatures/little-dragons-sea-69048 [accessed Nov 12, 2021]

24. Unity Technologies. Unity Asset Store. 2021. Available from: https://assetstore.unity.com/ [accessed Nov 12, 2021]

25. Unity Technologies. Unity Asset Store: Cute Pet. 2021. Available from: https://assetstore.unity.com/packages/3d/characters/animals/mammals/cute-pet-96976 [accessed Nov 12, 2021]

26. Sketchfab. Pony Luna 3D Model. 2021. Available from: https://sketchfab.com/3d-models/pony-luna-144fe93ac4f54651b25f753067e80d45 [accessed Nov 12, 2021]

27. Incompetech Inc. Royalty-Free Music by Kevin Macleod. 1997-2021. Available from: https://incompetech.com/music/royalty-free/music.html [accessed Nov 12, 2021]

28. Font F, Roma G, Serra X. Freesound technical demo. Proceedings of the 21st ACM international conference on Multimedia. ACM; 2013 Oct 21-25; Barcelona, Spain. New York: Association for computing machinery; 2013.