Using Software to Aid Vocabulary Delays Via Audio Stories - A Usability Test

Snezha Dimitrova

s2243652

Bachelor Thesis Psychology

Faculty of Behavioural, Management, and Social Sciences (BMS)

University of Twente

1st supervisor: DR. A.H. GIJLERS (HANNIE)

2nd supervisor: DR. A.M. VAN DIJK (ALIEKE)

February 2023

Abstract

This study aims to provide insight into the usability of a software tool that provides audio stories to young students to enhance their vocabulary skills without needing parental assistance. The software included audio stories and a word game modelled on the Peabody PPVT-III-NL test. Two versions of the app were created: passive and interactive. In the passive one, the story was presented in a linear order. In the interactive version, children were asked how the story should progress at different points. The aim was to increase a more engaging listing experience. Five Dutch children between the ages of 4 and 5 and a half were recruited, and the software was tested with them and their parents in the room. Results indicated that the children could navigate the software, but the instructions were insufficient for independent use. Additionally, issues were found with the length of the stories and the clarity of the pictorial representations of buttons. This is problematic as the target group could not read yet. The word game was well received, but minor issues were found with some images and the need for more randomization in the post-test. In conclusion, by improving the instructions and buttons and reducing the length of the stories, children should be able to use software to enhance their vocabulary skills without needing parental assistance.

Keywords: vocabulary development, audiobooks, early literacy software, interactive storytelling, language development, early literacy.

Introduction

Research has shown that there is a wide range of vocabulary size among children of the same age, with some children knowing fewer words than others (Fernald et al., Kleeck2013). This proves to be a problem as Segers and Vermeer (2008) have found that vocabulary knowledge predicts school success (Catts et al., 2003; Van Kleeck, 2008). In fact, delays in vocabulary often persist throughout a student's academic career (Biemiller and Slonim, 2001). This persistence can be explained by the fact that classroom instruction relies heavily on verbal and written communication, so the more words a child knows, the easier it is for them to acquire new knowledge (Lowry, n.d.). Reading and listening during classroom instruction require the child to recognize their meaning, also known as word comprehension (Catts et al., 2003).

The process of vocabulary development in young children can be influenced Hartby various factors, including the exposure to oral language and direct teaching of individual words and their meanings. Research has shown that the number of words a child is exposed to on a weekly basis significantly impacts their vocabulary knowledge (Farkas & Beron, 2004; Hargrave & Sénéchal, 2000; Hart & Risley, 1995). Hearing words in a concrete context assists children in the process of attaching labels to objects and understanding their words.

In particular, studies have indicated that a significant portion of a child's vocabulary is derived from the vocabulary of their parents, with socioeconomic status (SES) playing a role in determining the size of this vocabulary (Hart & Risley, 2003). One way to increase children's exposure to new words is book reading (Biemiller & Slonim, 2001). Stories present words in a specific context, and research findings illustrate that words in context are easier to learn (Nagy et al., 1985; Kuhn & Stahl, 1998; Segers & Vermeer, 2008; Sternberg, 1987). Furthermore, a read-aloud story exposes children to words and constructions they may not encounter in their oral environment (Casbergue & Harris, 1996). Therefore, stories can be used to expose children to new words while not being limited by their parents' vocabulary level. Unfortunately, access to reading material is also affected by the family's SES, as

income determines the ability to purchase print materials and literacy support (Linebarger, 2001). Even if low SES families are provided with reading material, it is not sure that parents will take the time to sit down and read with the kids (Linebarger, 2001).

A way to battle this issue is by using software to provide the literature material to the children. Additionally, to provide more literature material, the software makes the material easier to access as the software is always accessible either at school, at home or in the library (Díaz & Signes, 2014). With the current prevalence of smart devices and overall access to technology among young children, as reported by Van Kruistum and Van Steensel (2016), it is hypothesised that digital software could provide an effective method for increasing vocabulary without the need for parental involvement.

As the target group is still unable to read, there are two ways of delivering the stories to the target group - one is the form of written text, where the parents will have to read to the children or via audiobooks. Using audiobooks will eliminate the need for the parent to be involved in the listening, which aligns with the aim of this study. Audiobooks have been found to positively impact primary education learners as they improve their vocabulary and critical listening, build on prior knowledge and increase comprehension like any other literature format (Díaz & Signes, 2014). The benefits of using any variation of read-aloud stories have been found in multiple sources. According to Larson (2015), the benefits are the same for reading aloud by a non-interacting parent or teacher as for listening to pre-recorded stories in audiobooks. These findings were supported by Maher (2019), who did not find a significant variance between computer read-aloud and parents' reading on comprehension levels or focus (Segers et al., 2006). Therefore, both mediums could be used interchangeably to provide significantly similar results. Furthermore, audiobooks have the advantage of always being accessible either at school, at home or in the library (Díaz & Signes, 2014). In conclusion, an audiobook software format would provide the best results in delivering the stories.

In research, there are two story delivery types for increasing vocabulary, which are passive stories (PS) and interactive stories (IS) (Miller, 2004). Passive stories are the most

common type of stories where the storyline is singular, and the listener is a passive observer of the story (Miller, 2004). This is often the format in books, audiobooks, movies, and more, with a few exclusions when they are specifically designed to allow interactivity with the story. PS has been found to have a significant effect on improving children's oral language, meaning improving their pronunciation and adding new words to their vocabulary (Best, 2020; Biemiller & Slonim, 2001; Diaz & Signes, 2014; Casbergue & Harris, 1996; Hargrave & Sénéchal, 2000; Sternberg, 1987).

On the other hand, considerable research has found that most types of IS give similar results to PS on vocabulary acquisition (Penno et al., 2002; Segers & Vermeer, 2008; Smeets & Bus, 2012; Smeets & Bus, 2015; Verhallen et al., 2006). There are multiple ways of making a story interactive, such as adding elements like hotspots and multiple-choice vocabulary questions during the listening of the story (Smeets & Bus, 2012; Smeets & Bus, 2015). Another way is adding multimedia such as music, moving pictures, and sounds (Verhallen et al., 2006).

However, it still needs to be established that young children between the ages of 4 and 5 and a half are capable of using such a source to improve their vocabulary without the help of their parents. Usability issues could be if the child knows how to use a mobile device if they can understand the instructions given for each task and lack of interest that might lead to not using the software. Another usability issue could be caused due to complexity or number of tasks the child needs to complete. For this reason, it was chosen not to use added elements but rather to make the storyline interactive. This is modelled on the "Choose your own adventure" books as one of the oldest ways to make a story interactive (Adams et al., 2019). A typical story has a pearl structure where each event is followed by the next without breaks (Miller, 2004). Instead, the storyline can have a branching structure. In branching structures, the reader encounters forks in the path of the events, where they have to choose which path the story should take to continue. Once one path is selected, the story continues until another fork in the path is encountered (Miller, 2004). This type of story provides interactivity during the listening and autonomy to the listener over the story

development. Since branching structures as interactive stories have yet to be tested on vocabulary acquisition in young children, a PS version would also be included in the software.

Some tests will need to be added to the software to measure if vocabulary growth occurs during the audio stories. A game will be created and modelled to the Peabody PPVT-III-NL test to test for that. It is a worldwide used instrument to test vocabulary knowledge by measuring the understanding of spoken words (Pearson Clinical Assessment). The observer presents four images to the child and voices a word. Then the child has to choose the picture that best describes the spoken word where only one image is the correct answer.

In the present study, we want to assess if children of age 4 up to 5 and a half can use the software independently to improve their vocabulary and help prevent delays in their language development. The software will be designed and built as a low-fidelity prototype for that purpose. The software will contain two types of stories, passive and interactive, as we know from research that both positively affect vocabulary growth in children. Additionally, the software will incorporate a simple game through which the researchers can evaluate the children's learning of new words. Therefore, the research question is, "Can children aged between 4 and 5 and a half independently use software containing stories, interactivity elements and a game to increase their vocabulary?" Next, a detailed description of the application's design will be provided.

Methods

Research Design Overview

In this research, software was developed that offers both PS and IS that children can listen to. The tool also contained a test module that assessed the participants' knowledge of words before and after listening to those stories. The test module included words suited for the target group and related to the stories presented in the app. This paper presents the

results of a first evaluation with participants from the target group to gain insight into the software design's effectiveness and usability.

Participants

The study used a convenience sample. Children were recruited via the employees of a local company where the observer worked. Employees with children between four and five and a half years old were asked if their children would like to use the app. As compensation, children received a bag of sweets at the end of the study. Only the parents were made aware of the compensation beforehand. All parents were informed about the procedure of the sessions and the way experimental data would be handled. After agreeing with these policies, they signed a consent form. The consent form can be found in Appendix A.

In total, five children were able to participate between the ages of 4.0 and 5.5 years old, of which one girl and four boys. The children's average age was 5.28; the youngest was 5.0, and the oldest was 5.5. All the participants were native Dutch speakers and were born in the Netherlands. The children's parents were present during the study for translation, as the observer was not fluent in Dutch and for supervision purposes. All the communication between the researcher and the children was through their parents.

The number of participants should have been sufficient according to the general rules of usability testing presented by Nielsen (2000), where five participants were the recommended sample size. In his example, this has a 95% confidence level and a margin of error of ±18.5%, which was approximately 66.5% of 100% of the possible problems detected. Therefore, the minimum number of required participants for a usability test was set at 5 participants.

Materials

Stories

For the story material, permission to use copyright material was obtained from Dutch publisher Uitgeverij Moon. The book chosen was from a well-known Dutch children's books author Carry Slee. The book's title was "Het Grote Kwispelstaartjesboek", which was marked as suitable for children four years old and above by the publisher. Thai aligned with book classifications for the age group where each story contained no more than 1500 words, had an image every few pages and the topics were light-hearted (Murphy, 2019). The first eight stories from the book were selected for the experiment. Where four of them were made interactive, and the rest were kept in their original form. All stories were read and recorded by a native Dutch speaker. The length of the recorded stories varied from 7.04 to 11.24, with a mean of 9.38 minutes. In total, the PS length was 39,54 and the IS 33.35, with an average of 3.4 minutes between each break in the story.

As mentioned, the "Choose your adventure" type was used as a template for converting the stories into interactive ones. This type of book puts the reader in the shoes of the protagonist and invited the reader to choose how the story progressed. Usually, this was done when the reader was presented at multiple crossroads by making points in the story where the listener can choose how the protagonist proceeded (Adams et al., 2019). This resulted in a branching structure (see Figure 1) that could quickly become out of control and needed to be limited (Miller, 2004). To make it simple for children at age 4 to follow, the branching structure always contained just two choices, which eventually led to the same outcome as seen in Figure 2. The question attempted to be answered was "Is the length of the stories sufficient to keep the user engaged?".

Figure 1

General Branching Structure Example.

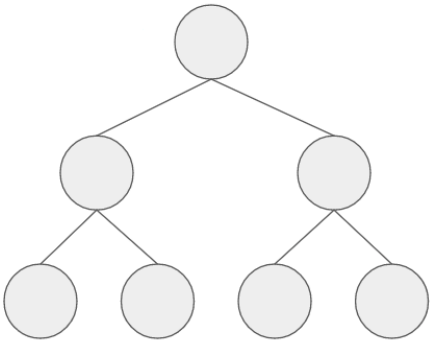
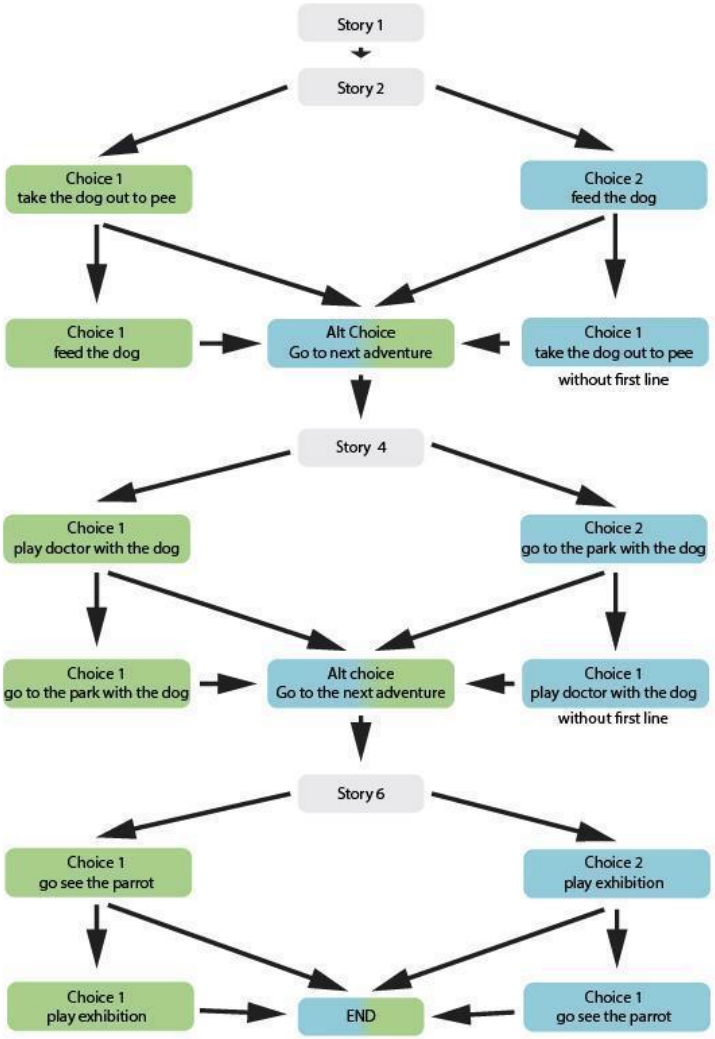


Figure 2

Customised Branching Structure for The Interactive Stories Found in The Software.



Word Game

The word game was used as a pre-and post-test to assess the pre-existing vocabulary knowledge and detect any growth in the vocabulary after exposure to the stories. The Peabody PPVT-III-NL test was chosen as a model for the game. The tool's reliability has sufficient coherence with the mean values between .89 and .9, and in the test-retest, the correlation was 0.94 (Pearson Clinical Assessment). The criteria for choosing the words from the stories to be tested was the Amsterdam Kindergarten Basic Vocabulary (BAK) list. The BAK was recommended by a local speech therapist practice. The list is a tool used by kindergarten teachers to expand children's vocabulary with 2000 words before they enter Group 3 (OCO, 2021). It was developed to battle language delays in Dutch children. BAK consists of words for group 1 and group 2. Each group has two lists of words; one is the minimum that needs to be known and an extension. The words chosen in this study are all from the Minimum list or the Expanded list for group 1, which is for ages 4 to 5. The eighth stories provided a sufficient number of words that could be tested for based on the BAK list, and the list of words created at the end contained 25 words, as seen in Table 1.

Table 1

Words Used in The Pre- and Post-Test.

Words in IS	Amounts of occurrence	Words in PS	Amounts of occurrence	Universal words	Amounts of occurrence
hamster	9	dame	15	schrikken	6
Leren	4	Knagen	5	park	9
plas	4	Knabbelen	4	Duw	73
buurvrouw	17	zolder	5	knuffel	19
papegaai	8	Halen	9	staart	16
omhoog	10	onder	11	paar	5
riem	11	winkel	9	Bang	7
ziek	6			achter	14

	liggen	13
	mand	5

The word game used in the current study consists of ten items per type of story (IS or PS). In the original test, an examiner reads-aloud a word and presented a paper with four illustrations, and the child was instructed to choose one that they thought described the word best. In the present study, we created a format in which the test was delivered through the app. The words chosen were all appropriate for group 1, which was for ages 4 to 5 and occurs at least four and more times in the stories. Most common connector words were excluded, such as “and”, “if”, and “all”, etc. and focus was kept on nouns, verbs and adjectives such as “fetch”, “anxious”, and “lying down”. As seen in Table 1, there were eight words found only in the IS, seven in PS and ten words across all stories. For the words to be distributed equally but randomly, the software took for the pre-test five words found in the type of story chosen to be listened to and drew another five from the words found in all stories. This resulted in having ten unique words to evaluate per story. This was done automatically at the beginning for each user, so every user has a different set of words in their pre-tests. The words assigned to each pre-test were later used in their respectful post-test with changed display order. The questions to be answered were “Can the users play the game by themselves”, “Were there any issues with the words?”, “Were there any issues with the images displayed for the answers?” and “Did the feedback images understand?”.

Smileyometer

To ascertain if the children liked the experience of using the software overall, a smileyometer was used. This provides a point of direct feedback from the participants. At the end of the observation, each child was presented with a Likert scale of five, represented with emojis (see Figure 3). As seen in Image 5, the far-left red icon was for the “I do not like it at all”, the orange one was “I did not like it”, the yellow icon was “I do not know”, the light green

“it was Okay”, and the dark green icon was for “I liked it a lot”. These later will be coded from one to five, from left to right, or from red to dark green. The aim of this material is to provide feedback outside of the observation as children at the target age group tend to be more unpredictable than adults in their body language and tend to be less verbal (Kercood & Banda, 2012).

Figure 3

Emojis Likert Scale.

Questions for parents

At the end of the experiment, the parents were asked an open question about their opinion of the experience for their children and an open question about their opinion on the software. This could provide further insight into how the parents perceived the possibility of their children being able to use the software on their own. The first question asked was, “What are the things you noticed your child likes, dislikes, had trouble with or found to be very easy?”. The second question asked to the parent was, “What is your opinion on the application?”.

Observation protocol

Additionally, the observer made use of an observation protocol during each session (see Appendix B). The observation protocol contained points of the software design in general, such as the quality of the instructions, usability of the sound button, images fit per word for the word game, the quality of the feedback, the length of the PS, the length for the IS, as well as quality of the choice buttons and end button during the interactive part of the stories. The quality instructions refer to if the child had any problems understanding what to

do after them or if they needed help. The usability of the sound button concerned if the children used the button and if they understood the purpose of the button. During the word games, images were chosen to describe the word and some filler images. It was essential to know if they were understandable enough and did not mislead the child in their answer. For the feedback, it was chosen to convey it through images. It was essential to check if the children understood the feedback.

Next, when it comes to the stories, it was essential to check if the length of the stories were not too long, which might lead to the premature ending of the session without completing the post-test. For the IS, we had more pages and, therefore, more elements to evaluate. In particular, the choice buttons for the next story to be listened to. The design relied on the images representing the stories to help the child to make a choice, not only on the action of the story but if they have listened to it before or not. Lastly, the end of the IS can be reached only if the child selected the end button when presented. If the child did not understand the meaning of that button, they would not have been able to end the listening when they wanted. The observation protocol can be found in Appendix B.

Software Design

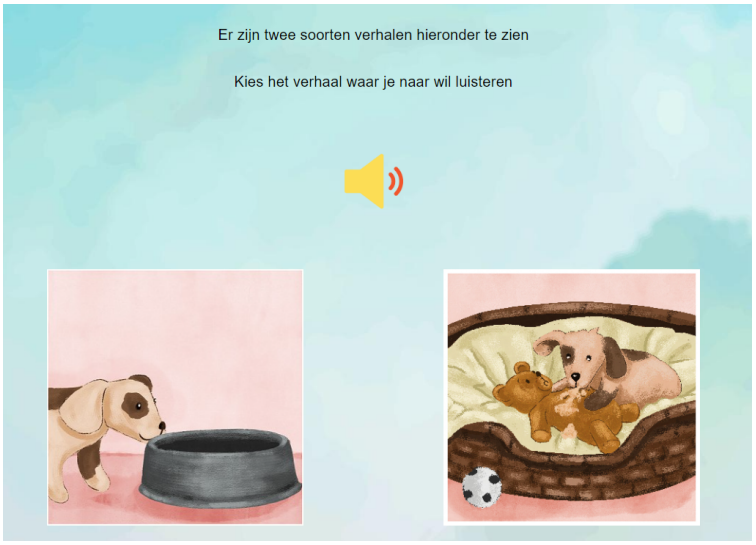
Other than the stories and the test, the software required other elements to make it possible for the child to engage and navigate the software. In that regard, there were four types of elements in the software - images, audio, buttons, and instructions. The images were required for the word game and visual engagement with the children during the stories. The audio was a crucial element of the software as it provided the children with both the stories, the word game and other instructions. Buttons allowed the child to provide their choices and navigate further in the software. Lastly, the instructions used were necessary to inform the child about the necessary actions they needed to take during the experiment. All images were copyrighted and used with permission or held by the researcher. All elements will be evaluated later to determine if they were sufficient to perform their intended function.

Images. The images in the software were used in the background, buttons and the word game. There was one general image for the background on each page. During the listening part of the stories, overlapping images depicted the actions taking place in the story that was playing. These story images were taken as a concept from the original book and drawn by an artist who relinquished their copyright to this research. The total number of story images was 19, for the PS 12, for the IS 6, and the first story had one picture. For the PS, while each story was played, one image from the story was displayed, and in the middle of the recording, the first image was replaced by a second one. For the IS, each of the three main stories was divided into two, so there were two images per part and four per main story.

The images used in buttons were for sound and choices in the IS. The image for the sound was a standard sound icon used most on digital devices. For the choices of stories, images from the stories were used in a scaled-down and cropped version. Examples can be found in Figure 4.

Figure 4

Page Where The User Needs to Select to Listen to Either IS or PS.



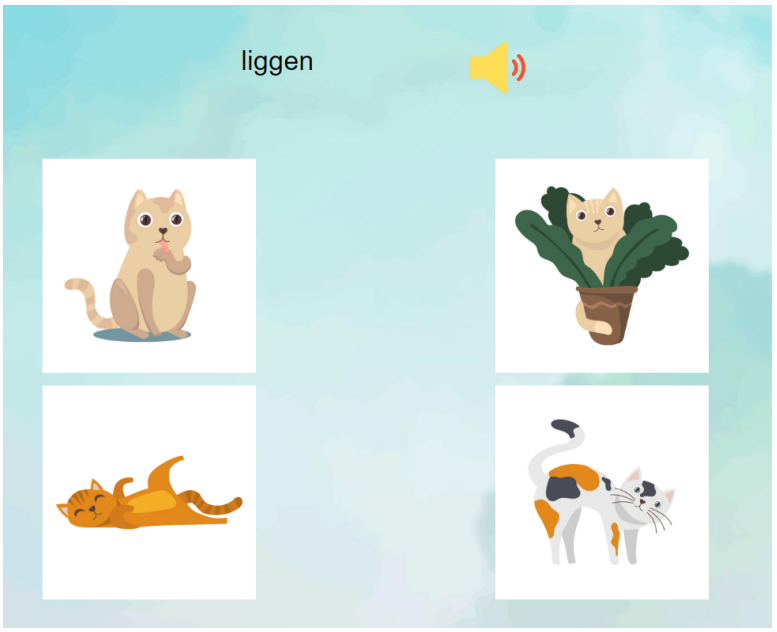
For example, story two had two branches where the main characters take their dog outside or feed it. Each choice was represented by a small square of the image that would

be displayed once they were chosen to be listened to. There was no right or wrong answer at this point.

On the other hand, the images in the word game were simple and coloured, representing the word in the test as best as possible. For reference, the type of images was similar to Peabody's test. All images used were copyrighted and used with permission. In total, the number of unique images used in the word game was 120, of which five were reused as filler images. In the word game, a feedback stage was added where another two images were used. One was a big red "X" for the wrong choice and a shooting confetti image for the correct choice. An example can be found in Figure 5. The images would be evaluated for the capability to add to attention retention when used in the pre-and post-test for accuracy and when used in buttons to represent the part of the stories to be chosen for listening.

Figure 5

Example Word Game Page.



Audio. Both the story-reading app and the test contained audio fragments. Audio fragments were utilised as the application's primary instruction method, as the target group cannot read yet. The different stories and instructions for the app were recorded using

podcast equipment to ensure optimal sound quality. A native Dutch student from the area of the target group narrated all audio fragments to account for any dialectical variations that might have existed in the targeted group. The same speaker recorded all stories and audio fragments to ensure consistency.

Instructions. The written text was primarily used for the parents whenever assistance was needed. This includes the consent form at the start page of the software. All instructions were written down on the choice pages and the words in the word game. There were three instructions made. The first was once the child lands on the homepage where they have to choose to listen to IS or PS. The instructions were as follow: “Er zijn twee soorten verhalen hieronder te zien. Kies het verhaal waar je naar wil luisteren.”. Which translates to English “There are two types of stories below. Choose the story you want to hear.”. Here we informed them there were two stories, and in the second sentence that, they will need to choose one. The second instruction was presented before the user started a word game. The instruction was as follows: “Nu ga je een aantal plaatjes zien. Kies het plaatje dat het best past bij het woord dat je hoort.”. Which translates to English: “Now you will see some pictures. Choose the picture that best matches the word you hear.”. Here the first sentence informed them that the next task included pictures, and the second sentence describes the game. The third instruction was displayed when the child has chosen to listen to IS. The instruction was to inform them every time they needed to make a choice about what the characters were to do next. The instruction was as follows: “Jij mag kiezen. Wat gaan Eefje en Mark als volgende doen?”. Which translated to English: “It is your choice. What should Eifje and Mark do next?”. The first sentence informed the child that they needed to make a choice, and the second used the main characters' names to describe that the choice was about them and what they would be doing next. The instructions provided in the software will be evaluated if they provide a sufficient understanding of the tasks to be completed. Questions aimed to be answered were “Were the instructions successful in relaying the task that needed to be completed by the user?” and “What issues were identified?”.

Buttons. Other elements found in the software were buttons. Examples of all can be found in Appendix C. The main buttons were for pausing and starting the audio recordings of the stories. A pause/start button was available during the full-length listening parts of all stories. Buttons also accompany all instructions with a sound icon if the user wanted to hear the instruction read-a-load again. The instruction recording would be played every time they clicked on the button. There was also a start button for the word game and an end button for the IS, which ended the story's flow. Another type of button was for the image choice for the word game and for choosing which story to listen to. There was a button to log in after putting a verification code at the beginning of the software. The included buttons were the minimum requirement for the user's autonomy over the software flow. Their design and usability will be under investigation. Questions aimed to be answered were "Can the user identify the button's function?", "Did the user need this function?", "Was the size of the button sufficient for selection?", "Was the end button recognizable as function", and "Did the choose button in IS distinguishable for the user?".

Software Procedure. The below overview provides a clear step-by-step task each child had to take during the experiment from their point of view. The software has a couple of sections - initial logins, initial choice page, initial story, passive and PS branches, and pre- and post-tests (see Appendix D).

The Initial logins encompass the start screen with the informed consent and the second page for the user's demographics such as age and gender. The demographics page was shown once per user. After the initial login pages, the software displayed the choice page. Here the user has to choose whether to listen to IS or PS (see Figure 4). Once the page loads, the instruction recording starts playing automatically, and once the user clicks on an image button, the software starts the initial story page. The initial story page plays the first story which introduces the main characters of the rest of the stories. It was the same for both PS and IS. The page has one image and a pause/play button. After the story was done, the instruction page for the pre-test was displayed.

On the instruction page for the pre-test, there was a written instruction and a sound button that could be pressed to listen to a recorded read-aloud of it. Under those, there was a start button. When the page was initially loaded, the read-aloud instruction was played automatically. Once the start button was pressed, the word page was displayed. On the word page, the word to be matched with its pictorial representation was written and played in recorded read-aloud. There was a sound button again for replaying the read-aloud. Under them were four square image buttons with possible choices for the user (see Figure 5). Once one was clicked, the feedback page was shown.

The feedback page contained the choice image the user selected, and additional images were displayed depending on if that was the correct image. This screen was displayed only for 10 seconds before the new word page was shown, and this was repeated until all ten words were matched. In the case of PS pages, after the pre-test, the initial page started with one image and a pause/play button was displayed. The recorded stories started to play, and all PS were played one after another without a break between them. In the end, the post-test instruction page was displayed.

In the case of IS pages, there was an introduction to each story, and then two choices were presented for the story to continue. During the introduction of each of the main three stories, the page contains only a background image and a pause/play button. The choice page was displayed when a fork in the story was encountered. The page has the same layout as the home page (see Appendix C). When the page started, the instruction audio recording started automatically. Once a choice was made, a story page was initiated with the audio recording, background image and a pause/play button.

While in the PS, the end was automatic, in the IS, the user chose an end. When an end became available, the choice image button on the right was replaced with an end button. Therefore, the user can choose if they want to go back and hear parts, they have not yet heard they can (see Appendix C). Once the end button was selected, the instruction page of the post-test was displayed. After completing the post-test, the user was reverted to the home page where they can choose to listen to the same story again or the other type.

Data Collection and Analysis

All observations were made in laboratory settings. Each child was led to a room clean of decoration equipped with a desk, chairs, a computer, and a camera. The camera was located above the child and beyond their view. The software was loaded to the initial page for the verification screen. During the observation, the parent remains in the room during the observation with the child and the observer. The parent was asked to sit next to their child in front of the computer. This was required in case the child could not use a mouse or had any questions. The parent was instructed beforehand not to help with their tasks, such as telling them the test's answer or explaining the task without the child asking for help. The observer put the participant number in the text field and put the age and gender of the child. Once the home page with the first choice was present, the observer did not take any actions until the end of the post-test. The exception is when the software runs into technical issues that need to be resolved for the child to continue participating. On average, the observation lasted 49.01 min, where the longest was 51.33 min, and the shortest was 45.89 min.

At the end of the study, the smileyometer was used first with the child and then the parent was asked the three open questions. In the end, before the child and their parent left, the child was given a present bag filled with sweets. The observation sessions were recorded with video and audio, which were made available to the observer if required a second observation in case something was missed.

The last data collected was from the software itself. As it collected information, such as how long it took to choose an image during the pre-and post-test, that allowed for further analysis. In the case of the word game results, it could be compared if there was a difference in the time needed to choose between the pre-and post-test. Any differences were not indicated if the child learned a word but if the child learned the answers through repetition or feedback. Furthermore, we could see if the child made fewer mistakes during the post-compared with the pre-test.

Additionally, it gathers the participants' demographics and which stories they chose to listen to. Especially when the child was listening to the IS, we could see their chosen path through the workflow of stories. In this case, selecting to listen to the same story indicated that the child did not recognise that they had listened to that story by the image for the choice or did not know what the end button represents.

Results

To keep the participants anonymous, they will be referred to with letters from A to E. The results will follow the observation protocol and present both the observer's observations and the parents' opinions. Participants A, B, and D chose to listen to the PS, and C and E chose the IS. The observation protocols can be found in Appendix B.

During the use of the software, observations were made on its parts' usability. Those under observation were the instructions given to the children, the stories, the integrated test before and after the stories, and the presented visuals and buttons. Lastly, it needs to be mentioned that all children were fidgeting, touching their faces and seemed distracted from an adult point of view. Unfortunately, this is typical behaviour for the target age group, which might indicate boredom, trying to maintain focus, and sometimes indicative of enjoyment (Kercood & Banda, 2012). Lastly it is important to mention that none of the user's knew prior to the test how to use a computer mouse, though two of the children did try and successfully used the mouse. Nevertheless, it is strongly recommended to use touch screen devices for the software in the future.

Observation protocol & Software

During the use of the software, observations were made on its parts' usability. Those under observation were the instructions given to the children, the stories, the integrated test before and after the stories, and the presented visuals and buttons.

Instructions and Buttons observations. During the use of the software, observations were made on the instructions and buttons' usability. Observations during the instruction phase suggested that children experienced difficulties operating the software independently. The children did not express their problems verbally but looked to their caregivers to request help. All parents reacted by paraphrasing the instruction and pointing to the screen to demonstrate the action they were talking about.

Instructions were offered at three different time points. The first part of the instruction was presented on the home page, where the child needed to choose to listen to either IS or PS. During this part of the instruction, participants A, B, and E had issues understanding what to do. After hearing the instruction, each looked at their parents and only did something further once their parents explained it further. The children were not asked what they did not understand, and the parents used verbal and physical instructions. The parents paraphrased the instructions given by the software and pointed at the screen where they needed to click to make a selection. Therefore, there needed to be more information to conclude if the children needed a better explanation or an example. The only conclusion that could be drawn was that the instructions were insufficient for the children to be able to work on their own.

The second instruction was presented before the pre-test to instruct children about the test. Here all participants except D needed help understanding. After hearing the instruction, all just clicked on the Start button, and none of them stopped to look at their parents first. However, when the first word appeared, the children either looked at their parents for directions or investigated what to do by hovering the mouse across the different elements. The delay in making the first choice was also visible in the data collected by the software. The average time each child took to choose the image fitting the word in the test was calculated (see Table 2). On average, each child took longer to respond to the first word presented, except for participant C. Where participant A took more than 72% more time, B took more than 54%, C took more than 9%, D took more than 47%, and E took more than 31%.

Furthermore, each child had a unique first word in the pre-test, so the time was not affected by one specific word. Since just one of the participants stopped at the instruction but seemed to need further instruction on the next screen, it could mean that they disregarded the instruction all together. Another explanation was that the instruction was not sufficient or in the correct place. This issue affected the pre-test results by changing the average time each child needed to answer a question but did not affect their capability to choose the correct answer. From a usability point of view, the instruction might have required an example as it is done in the original test this was modelled on.

Table 2

Time Spent on Matching Words with Images per Pre- and Post-Test per Participant.

Participant	Pre-test average choice time	Post-test average choice time	Post-test choice % of pre-post choice
A	12.600	4.110	32.62%
B	8.150	3.650	44.79%
C	6.420	3.460	53.89%
D	6.890	2.680	38.9%
E	6.980	4.330	62.03%

The third instruction was presented to children interacting with the IS version of the app during the listening session. Here children were asked to choose which story they wanted to listen to. While the requested action seemed similar to the action in the first instruction, participant E was unsure of what to do, and he looked up at his mother for more instructions which she gave him. The other participant who listened to the IS had no issues with the instruction. Since it was just two participants, it cannot be concluded with confidence that there were any issues with the third instruction.

The sound button was present during all instructions and the word game. Participants D and E indicated they needed to hear an instruction or word again. In this situation, the observer had to intervene and demonstrate the sound button's function. Which indicated the parent did not recognize the button's function. The rest of the participants did not need the button during the study, and no information was gathered about their understanding of it. While the function proved to be helpful, two of the participants needed help with the way the functionality was represented. There were no issues with selecting the button. These results indicated issues with the representation of the button's functionality and affected the usability of the sound button.

Passive Story and Interactive Story observations. During the use of the software, observations were made on the usability of the PS and IS, including the use of the choosing and end button in the IS. Observations of children interacting with the stories suggested that the length of the stories needs to be shortened. It was observed that all children had trouble focusing on the story. In the PS, this was indicated by one child that dozed off during the story, a child that kept asking when the story would end and a child that kept seeking distractions (trying to communicate with the parent etc.). For the PS participants, A and B needed help keeping their attention in the middle of the second story.

For the IS participant, C seemed focused until the last story. As mentioned above, the children had issues finding how to end the listening part and ended up listening to some parts multiple times. This was the case with participant C, who showed frustration. Participant E was tired from the beginning as they joined after swimming classes and was more anxious to leave and go home. While he was tired, he completed all tasks.

Towards the beginning of the third story, the observer noticed the child losing interest in listening. The child stopped looking at the screen and played with a paper in front of him. Overall, there was more attention towards the screen during the IS stories than the PS, but there was insufficient data to put the difference in numbers as the study recordings were not clear enough for eye tracking. However, from the overall observations, the children with PS

asked earlier about the remaining time left compared to IS. For the children to tell when the story is going to end without needing to ask, a progress indicator is needed. Furthermore, the length of the stories should be shortened or a break added.

For the IS, the chosen image and end buttons were investigated too. Both participants, C and E, had issues ending the stories with the end button and listened to the last stories twice. When presented with the end button, there would also be a button with a picture that led to a replay; they always chose the image. The observer carefully examined the facial and vocal expressions of the child to check whether the child wanted to replay the story or did not know how to end the story. When the child was irritated or disappointed, the researcher assumed the child did not wish to replay the story. This could be explained by their misunderstanding of what the button was or that it was a button. One of the parents also indicated that their child does not know how to read, and the button was unclear to them as it contains only text and no visual representation. Therefore, the end button was not suitable for the target group and needs to be replaced with a visual representation instead of text. While the choice buttons' images were also found insufficient in being distinguished as a previously made choice.

Pre- and post-test. In the pre-and post-test under investigation are the images chosen to represent the words and the feedback. Two words seemed to be unknown, or their pictures were ambiguous to the participants. Participants A and B did not know the meaning of "paar", while for participants D and E, the word "schrikken" was unknown. For the word "paar", none of the participants could guess the meaning of the word from the images. For the word "schrikken", both children tried the image of "bang". These results might indicate issues with the design choices for the images representing the words in the test.

Participant A had the word "bang", and while they did choose the correct image, they took significantly longer to make their choice. Participant A spent, on average, 12.600 seconds choosing an image per word, while for "bang", they took 21.249s. The problem was that the images for both "bang" and "schrikken" were used as possible answers, and this led

to confusion due to the similar meaning of the two words. This issue can be resolved by not using similar images representing similar words as possible answers within one question.

Another interesting word that was observed was “achter”. Only participants B and C had it in their test. While neither got it wrong, it was noticeable in the observation that they expressed hesitation. The backend data from the software shows that participant B spent 10.049s from his average of 8.150s, and participant C spent 17.766s from his average of 6.420s. Participant B did not spend that much time on the word, but participant C had a significant difference which could suggest issues with the image’s complexity or accuracy.

The following two words of interest were “buurvrouw” and “dame”. The word “buurvrouw” was presented only to participants C and E. While both got it correctly, there was some delay in the choice in the pre-test. For participant C, the delay was not much from, on average, 6.420s, he spent 8.176, while participant E spent almost double the time from normal - from on average 6.980s to 12.950s. As the word relates to the very narrow meaning of female neighbour, representing that concept proved difficult in the creation stage. Due to the small sample size, it cannot be concluded whether this would be an overall issue for the target group. For usability purposes, more attention might be needed when representing such words. The second word, “dame”, had a similar issue to the “buurvrouw”. Only participants A, B, and D had it. While participants B and D had no delays or incorrect answers, participant A got it wrong in the pre-test. From the observation, it was seen that participant B was hesitant. They moved their mouse between the images for the “dame” and the image for “queen”, while participant A pointed at the image “queen” and showed surprise when the feedback was negative. Here the issue might lie in having two images that could be interpreted as the word tested.

For the feedback screens, the overall response was positive. All children seemed to understand the images as wrong and correct. However, two of the parents also felt inclined to vocalise the feedback. Adding an audio element to the feedback might be beneficial.

Additionally, it was noticed during the observation and later confirmed from the software data collection that the children made choices faster in the post-test than in the

pre-test. As seen in Table 3 below, all participants spend approximately 55.58% less time on the post-test. While it was expected for the children to get better results in the post-test, this percentage might indicate that they could remember the positions of the answers or the correct answer. This indicates an issue with the design of the post-test.

Table 3

Descriptive Statistics for Mean Duration of Word-Image Choice for All Participants per Pre- and Post-Test.

Type	N	Mean
1	50	8.208
2	50	3.646

From the observation, it was also noticed this phenomenon. From the data collected by the software, it could be seen that each participant had a different improvement time (see Table 2). Participant A was 67% faster in the post-test average choice of image, participant B was 55% faster, participant C was 46% faster, participant D was 61% faster, and participant E was 37% faster. While the order of the words was randomised between the pre-and post-test, during the observation, it was visible that the children did not scan the images as they did during the pre-test, and they just selected the right image. The fastest choice was made by participant D twice, with a record of 0.3s, followed by participant C with 0.4s, and Participant A with 1.2s. This could indicate that the children either remembered the position of the correct answer or the image that was the right answer. While this does not directly affect the software's usability, it draws attention to the software's aim. For the software to be effective, the post-test needs to be improved concerning the design features.

To summarise, the users had no significant issues with playing the game that was the test. There were issues with some of the images chosen to be presented in the test, where there was more than one image that seemed close to the meaning of the word. Additionally,

issues with the post-test design were identified. The post-test needs to add randomisation of the answer images for each word or possibly have different images for the answers. The feedback images were comprehended, but since some parents did vocalise the feedback, an audio representation could be included.

Feedback Children and Parents

The feedback from the children was done via the smileyometer to allow the child to express their feedback non-verbally. This tool can compare the observations made on the software use and the children's thoughts. The feedback from the parents has the same intent to compare the observation made by the observer and the parent's observation.

From the children's Likert scale feedback, all participants chose the light green emoji, which was coded with 4. Even though Participant E seemed tired and fidgeting during the experiment, he verbally added that he liked the game at the end of the study. No other child gave verbal feedback. On the other hand, the parents all seemed optimistic about the software. The parent of participant E expressed how much her son liked the game even though he was tired. She also recommended changing the end button. The parent of participant A indicated that the pictures during the listening part were too boring. She further explained that it was harder to fall asleep when her children watched TV because their attention was fully taken. Suggested the use of more pictures. The parent of participant B also noticed that the child was bored during the listening part. She was surprised that her son did not know the word "paar". She further indicated that she liked the software and usually read to them at home. The parent of participant C mentioned that her child liked both the story and the test and recommended changing the end button to something more understandable. Lastly, the parent of participant D said the story was too long, but she still liked the idea behind the software.

To summarise, the children rated the software as likeable. Parents seemed to confirm the same. Further, according to the feedback the stories were interesting but possibly too

long or should include more images. Lastly, it was suggested for the end button to have pictorial representation.

In conclusion, the stories were liked, but there needed to be more clarity with the length and attention retention during listening, possibly by adding more images. The test was also liked, but there were issues with the images used in the test. It should be made so that there is not more than one image closely related to the meaning of the word. Moreover, the children seemed to either remember the position of the correct answer or its image in the test. Additional randomisation of the answers needs to be added, and possibly separate images for the answers in the post-test. The primary buttons needed during the story's use also seemed to have issues. The sound button was not clearly represented, and the end button requires a visual representation of its meaning as most of the target group cannot read it yet. Lastly, all instructions used in the software need to be improved by adding more explanations or/and examples.

Discussion

The study aimed at testing if software can be used to distribute audio stories to children to grow their vocabulary through independent work. The target group was 4- to 5-and-a-half-year-old Dutch children. The software's design was tested with the target group, and data about its usability was gathered through observations and feedback from the children and parents. The main problem that was interfering with the children's ability to work independently was the instructions provided by the software. They did not provide sufficient information for the children to complete the tasks independently. Next, all children required, at some point, information on how much longer the story was, which indicated the need for a progress indicator within the software. Finally, there were some issues with the word game. On some of the words, there was more than one image that was closely related to the word tested. Moreover, the children seemed to remember the correct answers from the pre-test, which affected the post-test results. Based on this information the answer to the research question "Can children aged between 4 and 5 and a half independently use software

containing stories, interactivity elements and a game to increase their vocabulary?" is that with a few modifications it should be possible for the children to use the software independently.

A lot of the issues could be affected by two factors, one was the presence of an adult in the room, and the other was the hour the study took place. According to Hanna et al. (1997), children best learn to use software through exploration. Their report states that preschool children would rather ask questions that explore when an adult was present. So, this could account for some of the times the children turned to their parents during the study for help instead of trying to figure it out themselves. Which also allows for the possibility that the children might have performed better with the software if they were alone. Furthermore, the study took place in the evenings for all participants, which could also influence their desire to explore on their own or diminish their interest in the software (Hanna et al., 1997). In this case two participants had problems, such as one child almost falling asleep, and the other wanted to go home as soon as possible.

A few factors might have caused the participants' issues with the instructions. The results found that all the children had indicated at least once that they did not know how to proceed after the instruction by looking up to their parents. One possible explanation is that the instructions were not elaborate or detailed enough for the children to understand. Since the parents did help their children by rephrasing the instructions, they also gave an example with their hands. This opens the possibility that the children might have needed an example and not necessarily the paraphrased instruction. In the study of Kähkönen and Ovaska (2006), children best responded to instructions implemented in the software containing verbal instructions and animated or pictorial examples of the task (Rodriguez & Polson, 2004). Martin and Smith-Jackson (2005) specifically found that this was true when the steps of the tasks were represented visually, and written instructions were not necessary for children. Considering this, a recommendation for the future was that all the instructions include an example of the task to be done verbally or with animation. Since the parents did both to help their children with the instruction, it cannot be concluded with certainty which or

both were needed. Adding to that is the fact that the parent was in a room which could have made the children more comfortable to ask first before they continued to find how to do the task (Hanna et al., 1997), which could indicate that there might have been no issue with the instruction in the first place.

In the case of the current software, there are multiple ways an example can be incorporated with the instructions. For the start page, it could be as simple as a hand cursor appearing and moving across the two options for stories. An example video on the instruction page should be included for the test. In the instructional manual of the PPVT-III-NL test, the linguist has to first play a fake round with the children until they are sure the child understands the task. That was done by the linguist showing the child four pictures on paper and giving verbal instructions combined with pointing with a finger at the pictures as choices. The same can be done in the software through a video showing the process of choosing an image that might fit the word heard, including verbal instruction and a cursor showing making a choice. Lastly, for the IS choice pages, to the current instruction, it can be added to the verbal instructions what the two choices are. And when a choice was said, the picture button was highlighted.

Regarding the buttons, there are issues with representing their functionality. While the sound button was needed only by two participants, neither seemed to know where to click to hear the instructions again. This could be attributed to either the button visual representation or the fact that the parents were in the room. In the study of Kähkönen and Ovaska (2006), children tended to find the button's functionality through exploration of the software or by mistake. Since the sound button has an important function, further instruction might need to be added at the beginning to introduce the button's functionality or to improve its visual representation.

Meanwhile, there were evident issues with the end button in the IS. When faced with the option to continue listening or end the listening with an image of a story and an end button, the participants always chose the image. First, the end button did not have a visual but only a textual representation. As found by Martin and Smith-Jackson (2005), young

children usually cannot read and would ignore textual information. This could explain the participant's preference for the image choice instead of the end button in the IS. As mentioned before, visual representation has the best results. Therefore, an image representing the finish line might resolve the issue. Of course, since only two participants chose IS, there was no sufficient data to give more information about using the end button.

When it comes to the issues with the stories, they might lie in the length of the story but also in the engagement of the child during the listening. Minimal images were playing during the stories, and no actions were required but listening, which could be the cause of the loss of interest so early in the stories. Also, the higher level of attention during the IS could be due to the child having more to do and more perceived autonomy. In total, all the children wanted to know when it would finish as they had no indication of progress. Therefore, there are multiple ways to resolve the issues. One possible way is by displaying the progress of the story to the children. This can be done by using a progress indicator displayed during the listening part. A progress indicator informs the user the remaining time until the next action to be taken (Sherwin, 2014). It reduces the user's uncertainty and perception of time and gives the user something to look at while listening (Sherwin, 2014). By adding a progress indicator it might stop the children from asking about the end and might affect the amount of eye contact with the screen. Therefore, it would be advisable to add such an indicator during the listening part.

In the original design, pictures were added during the listening part to provide some visual stimuli during the listening. If more attention was required for the screen, it would seem an increase in visual stimuli might be required. The meta-analysis of Takacs et al. (2015) found that technology-based storybooks are more effective when multimedia was added. However, this was not the case for stories including both multimedia and interactive elements. By multimedia, this study refers to animated pictures, music, and sound effects. For example, in their research Mutalib et al. (2011) used a lot of multimedia and autonomy elements in their interactive storytelling software, which led to positive engagement between the participants and their software. They added a narrator with a video or multiple changing

pictures to the story. Therefore, as an example, this study software could use either more images during the listening part or video to attract more attention to the screen.

In contrast, if the issue was indeed the length, then breaking the stories into smaller portions might be a solution. The research of Glenn (1978) found that the longer the story, the less it was recalled by the children. The theory of using an interactive story was to give more to do during listening and attract the child's attention to important elements. While shortening, the stories used might also improve the recall of the elements that were targeted to be learned.

The drawback of using any pre-test was that it can influence the post-test results, also known as a pre-test effect (James & Storm, 2019). This ensures internal but not external validity (Kumari, 2013). The transfer of knowledge from the pre-test to the post-test was almost always present, even in the cases where there was no feedback on whether the answers given in the pre-test were correct (Pan & Sana, 2021). Though it has to be mentioned that this does not affect the learning of information not mentioned in the pre-test (James & Storm, 2019). If the goal was to overcome the pre-test effect, it could be used the Solomon Four Test which adds another two control groups (Kumari, 2013). As in the current case, the problem was that the children might be remembering the position of the correct answers or the correct answers' image. A simpler solution would be to change the positions of the answers for each word in the post-test compared to the pre-test. This will prevent the children from going for the previous position of the correct answer. To eliminate the remembering of the image, a different set of images can be used in the post-test bearing the same meaning but looking different enough.

Additionally, a limitation of the study was the use of the low-fidelity prototype for the test. The developed software did not allow for independent work by the participants, as it gave a lot of errors that needed to be corrected by the observer. Such issues were that the stories would replay infinitely and not move further in the software flow. Furthermore, the software will not allow for interruptions in the sessions during the listening parts, which could also be a drawback for independent work with the software. As mentioned before, the study

was done in the evenings, which also seemed to contribute to the children's attention, patience and energy levels.

Moreover, since the software had to be used in a lab, the limitations for participation increased as now more resources were required. This meant that the study had to be conducted one participant at a time, requiring a room where to be conducted, an observer and one parent of the participants to be present. The presence of all the adults also seemed to affect the children's behaviour as, at their age, their attention span, motivation to please adults, and ability to adjust to strange surroundings and new people change rapidly (Hanna et al., 1997).

Based on the result above, it could be concluded that it might be possible if a few changes are made to the design that children of the target group will be able to use the software independently. In summary, changes are needed to the instruction used in the software to describe the action the child needs to take. The software might not be used for independent work without making them more elaborate and including examples. Next, shortening the stories and adding more multimedia during the listening would increase their attention. Adding a progress indicator during the listening part could also reduce the need for assistance. Moreover, the word game needs to ensure not to have more than one image closely describing the word tested so the measurement of the progress remains accurate. Finally, additional randomisation is required for the post-test answers as children seem to remember the images and positions of the correct answers. Overall, the software can serve as a groundwork for developing tools children can use independently to improve their vocabulary.

References

Adams, D., Beckelhymer, D., & Marr, A. (2019). Choose Your Own Adventure: An Analysis of Interactive Gamebooks Using Graph Theory. *Journal of Humanistic Mathematics, 9*(2), 44-59. https://scholarship.claremont.edu/jhm/vol9/iss2/5/

Best, E. (2020). Audiobooks and literacy A rapid review of the literature. *National Literacy Trust (United Kingdom)*. https://eric.ed.gov/?id=ED607775

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: evidence for a common sequence of vocabulary acquisition. *Journal Educational Psychology, 93*, 498–520. http://dx.doi.org/10.1037/0022-0663.93.3.498

Casbergue, R. M., & Harris. K. (1996). Listening and Literacy: Audiobooks in the Reading Program. *Reading Horizons: A Journal of Literacy and Language Arts, 37*(1). https://scholarworks.wmich.edu/reading_horizons/vol37/iss1/4

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping Poor Readers on the Basis of Individual Differences in Reading-Related Abilities. *Journal of Learning Disabilities, 36*(2), 151–164. https://doi.org/10.1177/002221940303600208

Díaz, M. A., & Signes, C.G. (2014). Audiobooks: improving fluency and instilling literary skills and education for development. Junta de Extremadura: *Consejería de Educación, 20*(1). http://hdl.handle.net/10662/4523

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research, 33*(3), 464–497. https://doi.org/10.1016/j.ssresearch.2003.08.001

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Glenn, C.G. (1978). The role of episodic structure and of story length in children's recall of simple stories. *Journal of Verbal Learning and Verbal Behavior, 17*(2), 229-247. https://doi.org/10.1016/S0022-5371(78)90164-0

Hanna, L., Risden, K., & Alexander, K. (1997). Guidelines for usability testing with children. *Interactions, 4*(5), 9–14. https://doi.org/10.1145/264044.264045

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1), 75–90. https://doi.org/10.1016/S0885-2006(99)00038-1

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. P.H. Brookes. https://psycnet.apa.org/record/1995-98021-000

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap. American Educator, 27, 4-9.

James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested? *Journal of Experimental Psychology: Applied, 25*(4), 576–587. https://doi.org/10.1037/xap0000231

Kähkönen, N., & Ovaska, S. (2006). Initial observations on children and online instructions. *IDC '06: Proceedings of the 2006 conference on Interaction design and children*, 93–96. https://doi.org/10.1145/1139073.1139098

Kercood, S., & Banda, D. R. (2012). The Effects of Added Physical Activity on Performance during a Listening Comprehension Task for Students with and without Attention Problems. *International Journal of Applied Educational Studies, 13*(1), 19-32.

Kleeck, A.V. (2008), Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychol. Schs., 45*, 627-643. https://doi.org/10.1002/pits.20314

Kruistum, C. van, & Steensel, R. van (2016). Young children (0-8) and digital technology: A qualitative exploratory study (Dutch National report). http://dx.doi.org/10.13140/RG.2.1.2683.4329

Kuhn, M. R., & Stahl, S. A. (1998). Teaching Children to Learn Word Meanings from Context: A Synthesis and Some Questions. *Journal of Literacy Research, 30*(1), 119–138. https://doi.org/10.1080/10862969809547983

Kumari, P.L. (2013). Significance of Solomon four group pretest-posttest method in True Experimental Research- A Study. *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS), 5*(2), 51-58. https://www.academia.edu/5307948/Significance_of_Solomon_four_group_pretest_posttest_method_in_True_Experimental_Research_A_Study

Larson, L.C. (2015). E-Books and Audiobooks: Extending the Digital Reading Experience. *Reading Teacher, 69*(2), 169-177. https://doi.org/10.1002/trtr.1371

Linebarger, D. L. (2001). Bookreview: beginning literacy with language: young children learning at home and school D. K. Dickinson & P.O. Tabors (Eds.), baltimore: brookes. *Topics in Early Childhood Special Education*, *21*(3), 188–191. https://doi.org/10.1177/027112140102100308

Lowry, L. (n.d.). *Build Your Child's Vocabulary.* http://www.hanen.org/helpful-info/articles/build-your-childs-vocabulary.aspx

Maher, C. (2019). Enhancing Literacy Skills Using Audiobooks Comprehension and Enjoyment of Text in the Second Grade Classroom [Master's thesis, Tampere University of Applied Sciences]. Tampere University of Applied Sciences Theses Open Collection. https://urn.fi/URN:NBN:fi:amk-2019060515102

Martin, C. V., & Smith-Jackson, T. L. (2005). Analysis of Assembly Instructions for Children: The Good, the Bad, and the Ugly. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(18), 1729–1733. https://doi.org/10.1177/154193120504901811

Miller, C.H. (2004). Structure in interactive media. *In Digital storytelling a creator's guide to interactive entertainment* (pp. 119-135). Elsevier.

Murphy, J. (2019). *Children's Books: Industry Standards and Resources*. https://blog.ung.edu/press/dlf-pres-2019/

Mutalib, A. A., Aziz, N., & Shaffiei, Z. A. (2011). Digital storytelling makes reading fun and entertaining. *International Journal of Computer Applications, 18*(1), 20-26. https://www.ijcaonline.org/archives/volume18/number1/2248-2878

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly, 20*(2), 233-253. https://doi.org/10.2307/747758

Nielsen, J. (2000, March 18). *Why You Only Need to Test with 5 Users.* https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

OCO (2021, 23 September). *Basiswoordenlijst Amsterdamse Kleuters (BAK).* https://www.onderwijsconsument.nl/presentatie-basiswoordenlijst/

Pan, S. C., & Sana, F. (2021). Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice. *Journal of Experimental Psychology: Applied, 27*(2), 237–257. https://doi.org/10.1037/xap0000345

Pearson Clinical Assessment (n.d.). PPVT-III-NL | Peabody Picture Vocabulary Test-III-NL. https://www.pearsonclinical.nl/ppvt-iii-nl-peabody-picture-vocabulary-test

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23–33. https://doi.org/10.1037/0022-0663.94.1.23

Rodriguez, M. A., & Polson, P. G. (2004). Creating Usable Wordless Instructions for Performing Complex One-Time Tasks: Effects of Violating the Rules. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(6), 741–745. https://doi.org/10.1177/154193120404800602

Segers, E., Nooijen, M., & de Moor, J. (2006) Computer vocabulary training in kindergarten children with special needs. *International Journal of Rehabilitation Research, 29*(4), 343-345. https://doi.org/10.1097/mrr.0b013e328010f4e0

Segers, E., & Vermeer, A. (2008). Vocabulary learning by computer in Kindergarten: the possibilities of interactive vocabulary books. In O. Saracho & B. Spodek (Eds.), *Contemporary Perspectives on Science and Technology in Early Childhood Education*

(pp.149-166). Information Age Publishing. https://www.researchgate.net/publication/293488734_Vocabulary_learning_by_computer_in_Kindergarten_the_possibilities_of_interactive_vocabulary_books

Sherwin, K. (2014, 26 October). *Progress Indicators Make a Slow System Less Insufferable.* https://www.nngroup.com/articles/progress-indicators/#:~:text=Progress%20indicators%20tell%20users%20

Smeets, D.J., & Bus, A.G. (2012). Interactive electronic storybooks for kindergartners to promote vocabulary growth. *J Exp Child Psychol., 112*(1), 36-55. https://doi.org/10.1016/j.jecp.2011.12.003

Smeets, D. J. H., & Bus, A. G. (2015). The interactive animated e-book as a word learning device for kindergartners. *Applied Psycholinguistics, 36*(4), 899–920. https://doi.org/10.1017/S0142716413000556

Sternberg, R. J. (1987). Most Vocabulary Is Learned from Context. In M. McKeown, & M. E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 89-106). Hillside, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781315802336

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and Pitfalls of Multimedia and Interactive Features in Technology-Enhanced Storybooks: A Meta-Analysis. *Review of Educational Research, 85*(4), 698–739. https://doi.org/10.3102/0034654314566989

Verhallen, M. J. A. J., Bus, A. G., & de Jong, M. T. (2006). The promise of multimedia stories for kindergarten children at risk. J*ournal of Educational Psychology, 98*(2), 410–419. https://doi.org/10.1037/0022-0663.98.2.410

Appendices

Appendix A

Informed Consent

Consent Form for Effectiveness of interactive storytelling on early literacy vocabulary acquisition

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated 16th of December 2022, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
I consent voluntarily for my child to be a participant in this study and understand that my child can refuse to answer questions and I can withdraw them from the study at any time, without having to give a reason.	☐	☐
I understand that taking part in the study involves my child listening to audio recorded stories and play a word game where they match a word they hear with an image.	☐	☐
I understand that information I provide will be used only for the purposes of a bachelor thesis project at University of Twente.	☐	☐
I understand that personal information collected about my child that can identify them, such as their age and gender, will not be shared beyond the study team.	☐	☐
I agree my child and me to be audio and video recorded.	☐	☐

Signatures

_____ _____ _____

Name of participant [printed] Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Snezha Dimitrova _____ _____

Researcher name [printed] Signature Date

Study contact details for further information: Snezha Dimitrova
s.dimitrova@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl

UNIVERSITY OF TWENTE.

Appendix B

Observation Protocol

General	Child question	Parent intervention	Child's behaviour
Instructions			
Sound Button			
Test Images			
Test Feedback images			
Passive story			
Length stories			
Interactive Story			
Length stories			
Choice button			
End button			

Appendix C

Images Design of Software

Figure C1

Initial Screen with Consent Form.

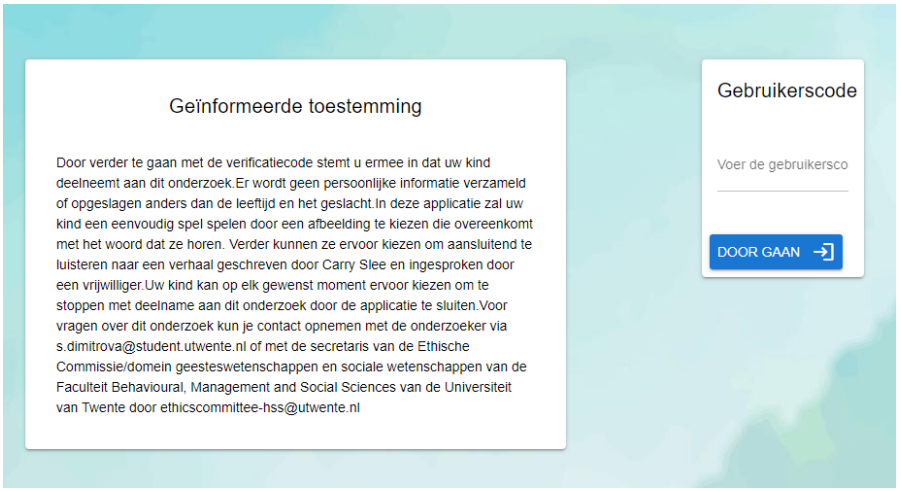


Figure C2

Demographics Page.

Figure C3

Choice Page Between IS and PS.

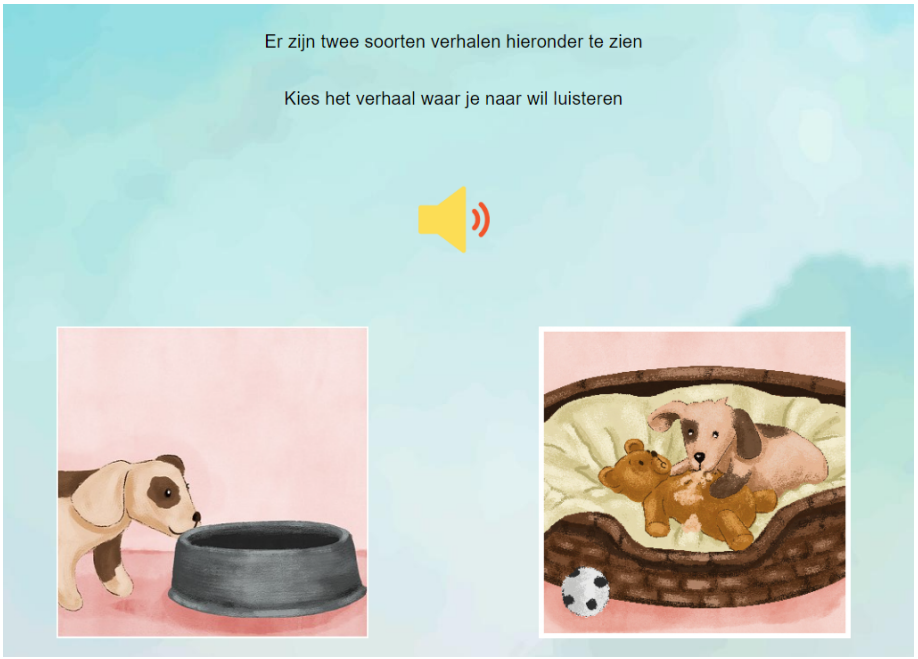


Figure C4

Instructions on How to Play The Word Game.

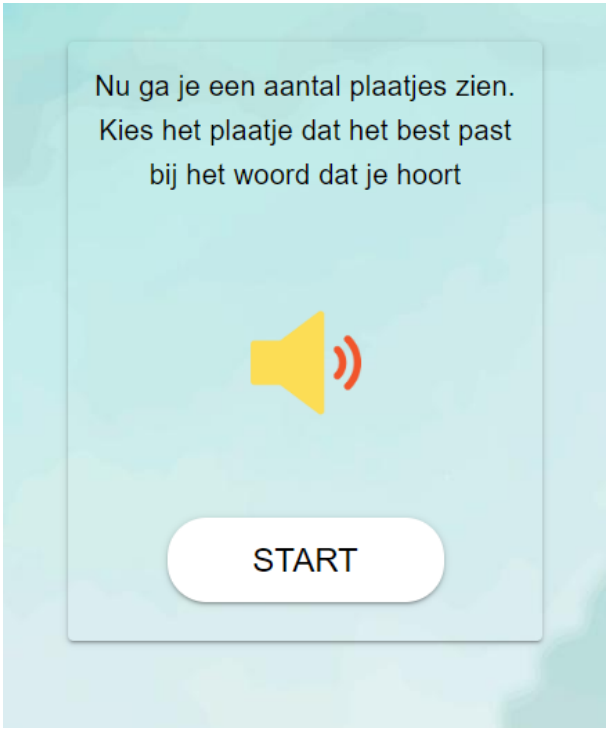


Figure C5

Example Page of The Listening Parts For Both IS and PS.

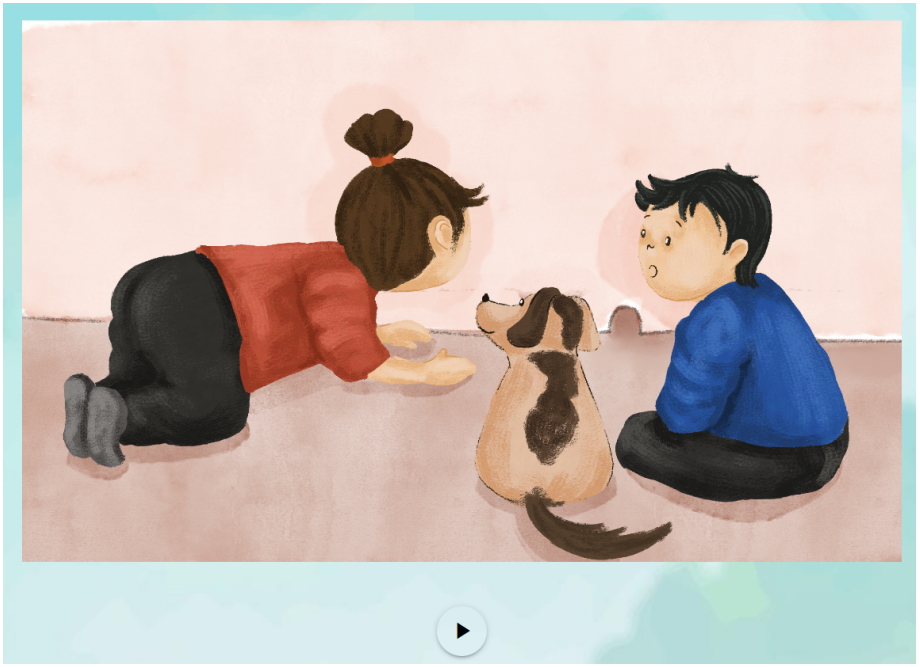


Figure C6

Example Word Game Page.

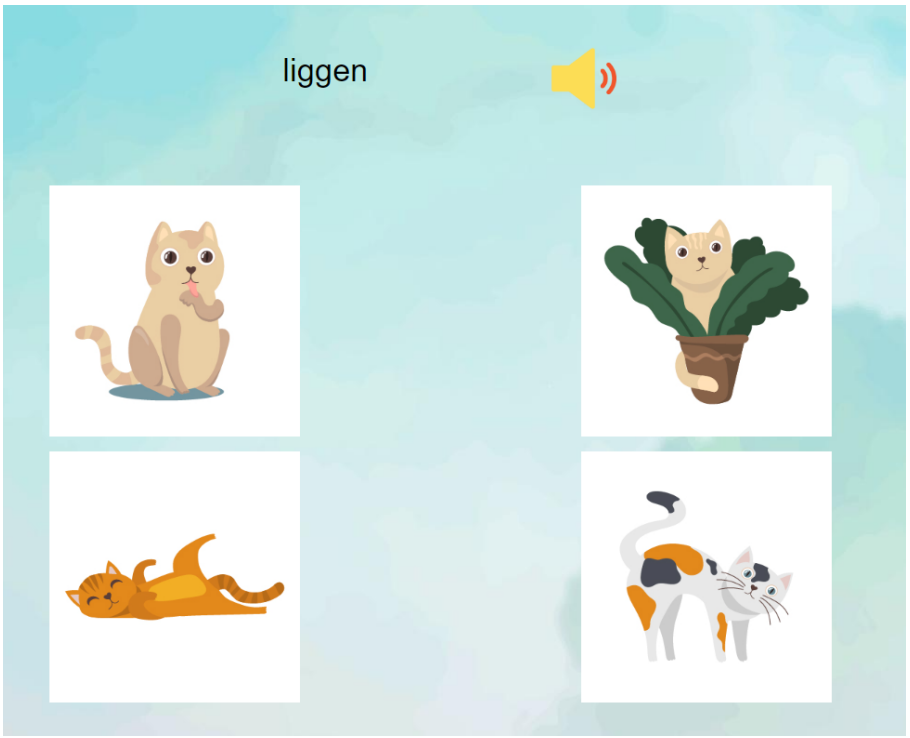


Figure C7

Example of a Positive Feedback Page.

Figure C8

Example of Negative Feedback.

Appendix D

Flow Charts Used in The Development of The Software

Figure D1

Flow chart of The Software Application.

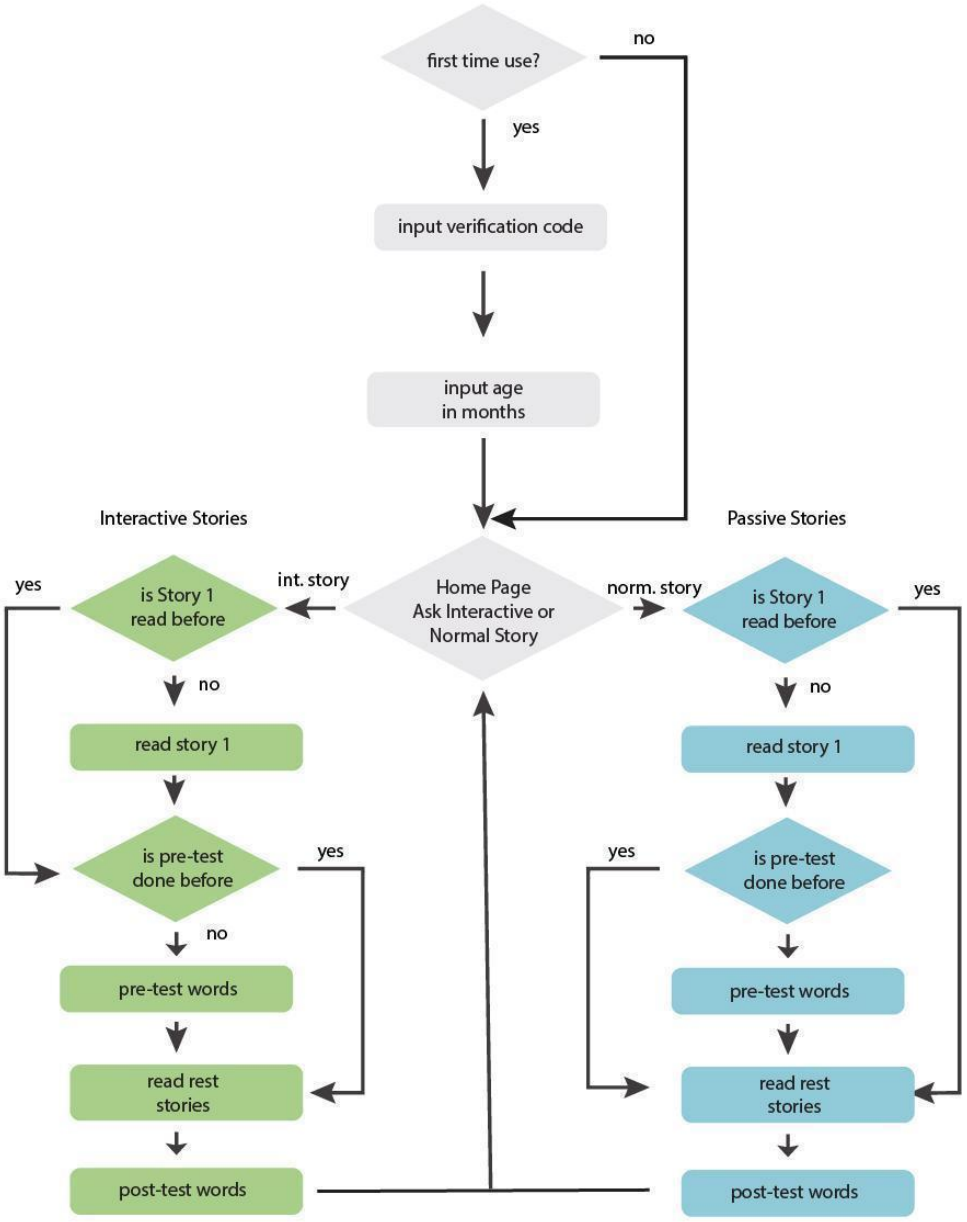


Figure D2

Flow Chart for The Word Game.

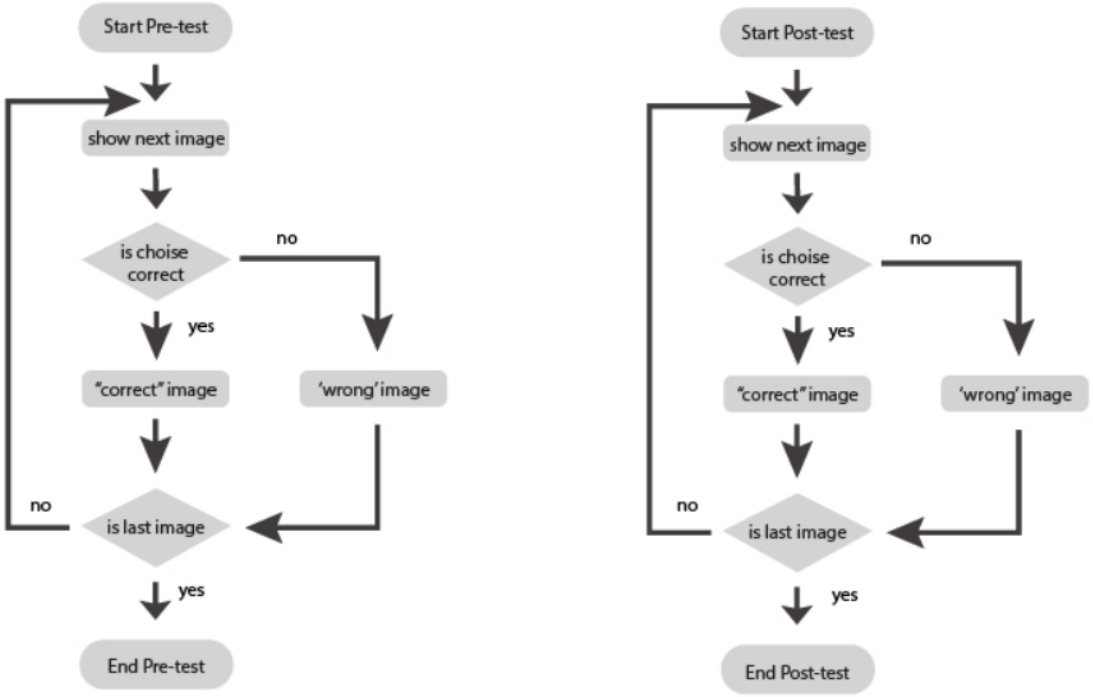
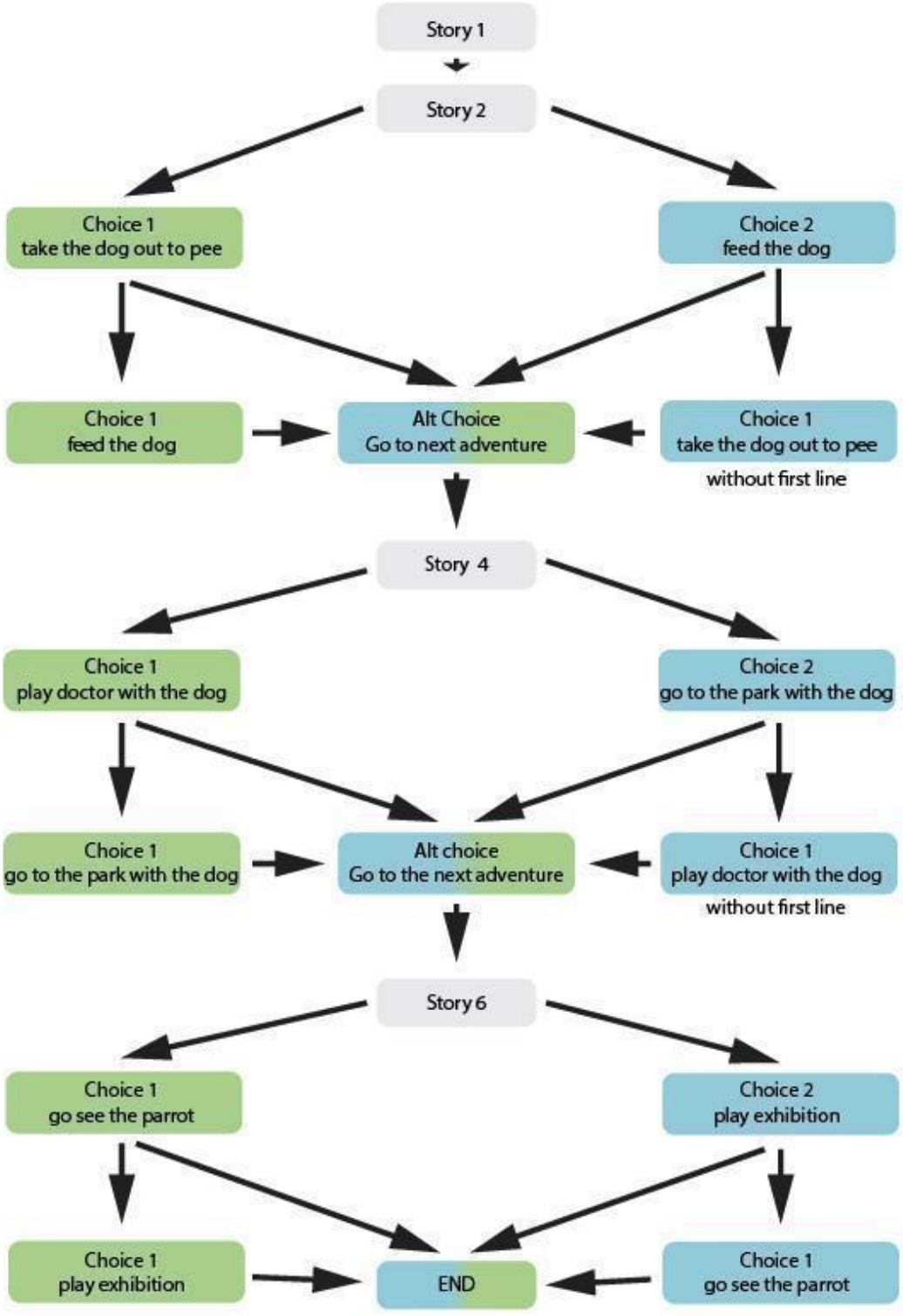


Figure D3

Customised Branching Structure for The Interactive Stories Found in The Software.


Story 1
Story 2
Choice 1
take the dog out to pee
Choice 2
feed the dog
Choice 1
feed the dog
Alt Choice
Go to next adventure
Choice 1
take the dog out to pee
without first line
Story 4
Choice 1
play doctor with the dog
Choice 2
go to the park with the dog
Choice 1
go to the park with the dog
Alt choice
Go to the next adventure
Choice 1
play doctor with the dog
without first line
Story 6
Choice 1
go see the parrot
Choice 2
play exhibition
Choice 1
play exhibition
END
Choice 1
go see the parrot