

RADBOUD UNIVERSITY NIJMEGEN

# Learning through play

---

A study on the effectiveness of the Noplica  
playhouse game “Dancefloor”

**Master Thesis**

**Elea Thijssen-Kolkman**

**S4261836**

**December 2020**

**Primary supervisor: Prof. dr. Paula Fikkert**

**Second reader: Laura Hahn**

## Acknowledgements

I would like to give my thanks to: Paula, for her trust, patience and inspiration. To Iris, Erwin, Greetje, Rianne, Jill, Annemiek, and all the children at OBS de Bloemberg for being so kind and helpful. To all my parents for their love, support and interest. To Julia, for her helpful comments and mental support. To Lucifer coffee roasters for providing the best roast and a peaceful work space. And thank you Theo, for making everything possible. I dedicate this thesis to you, knowing that everything will always be all right in the end, and that you are there besides me to guide me through the rough patches. Finally, Nora, thank you for brightening my every day.

## Table of contents

Acknowledgements .....	I
Table of contents.....	II
Abstract .....	1
1 Introduction.....	2
2 Method.....	6
2.1 Participants.....	6
2.2 Materials.....	6
2.2.1 The Noplica language playhouse.....	6
2.2.2 Vocabulary tests .....	9
2.3 Design.....	14
2.4 Procedures.....	15
2.4.1 Vocabulary pre-tests .....	15
2.4.2 Play sessions .....	16
2.4.3 Vocabulary post-tests.....	17
2.5 Apparatus .....	17
2.6 Analysis.....	18
3 Results .....	19
3.1 Descriptive statistics.....	19
3.1.1 The play session observation checklists .....	19
3.1.2 The classroom questionnaires.....	20
3.1.3 The vocabulary test scores .....	20
3.2 Exploring assumptions.....	22
3.3 Running the MANOVA.....	22
4 Discussion .....	23
4.1 Experiment results explanation.....	23
4.2 Play session observations.....	24
4.3 Suggestions for improvement of materials and procedure .....	25
4.4 Suggestions for other future research .....	26
5 Conclusions.....	28
6 References.....	29
7 Appendices .....	30
7.1 Appendix A: Dancefloor content .....	30
7.2 Appendix B: Granny stories .....	32

7.3	Appendix C: Vocabulary perception test.....	34
7.4	Appendix D: Scoring sheets .....	37
7.5	Appendix E: Vocabulary production test.....	39
7.6	Appendix F: Test item selection brainstorm .....	40
7.7	Appendix G: Consent Forms .....	44
7.8	Appendix H: Observation Checklist .....	47
7.9	Appendix I: Colouring pages.....	48
7.10	Appendix J: Classroom questionnaire .....	50
7.11	Appendix K: Overlapping words .....	51
7.12	Appendix L: Playhouse upgrades.....	52

## Abstract

For various reasons, children can start elementary school with a setback in linguistic development. One main cause is when they or their parents originate from a different country, and therefore do not speak the language (yet). For these children, their educational journey is exceedingly difficult, since they need to learn the language in order to understand the curriculum. But teaching them language when they might not understand the instructions and explanations is a problem. One way to solve this issue, is to look at different methods of learning. When examining and combining the different factors that are shown to facilitate language learning, **learning through play** is found to be a very effective method. This paper reports a pilot study to investigate the effectiveness of the Noplica language playhouse, that was specifically designed to facilitate language learning through play. One game in particular is chosen: the Dancefloor game, since it combines and balances out most of the facilitating factors. It is designed to stimulate vocabulary learning. The experiment consists of a two-part vocabulary test, looking at both perceptive and productive vocabulary, and a series of play sessions. The test is administered both before and after a period of four weeks, within which the participating children were taken to the playhouse bi-weekly, for a total of eight play sessions, lasting fifteen minutes each. Half the children played the Dancefloor game, and the other half played a different game, in order to single out the effectiveness of the Dancefloor game while keeping all other factors as comparable as possible. The improvement scores on the vocabulary tests were calculated and compared between groups. Even though there was a large difference, due to individual variation and small group sizes this was found to be insignificant. The materials and procedures used require some adjustments, but the overall method worked quite well, and the large effect sizes found in the analysis provide hope that a future larger-scale study will yield a significant effect.

# 1 Introduction

Not all children start at school with equal opportunities. They can come from varying backgrounds, which can have a significant influence on their pre-school development, especially when it comes to linguistic development. Some children speak one language, other children might know more. Some children have large vocabularies, others have smaller ones. Most children grow up in the country where they are going to attend school, and thus are likely to speak the language used at school, but others might have only recently moved to a new country and do not speak the school language at all. This initial setback is tough, because not understanding the curriculum makes it harder to keep up, and this inequality could be maintained throughout life.

In the Netherlands, immigration has been increasing (on average) since about 2007, and the total immigration for 2019 was 267.738 persons (Centraal Bureau voor de Statistiek, 2020). Most of these immigrants do not speak Dutch, and for their children, this means that going to school will require them to learn Dutch, before they can start working on the rest of the curriculum. For some children, this can be more difficult than others. There are various contributing factors, such as age of acquisition or linguistic distance. Another important factor that has been related to language learning is Socioeconomic Status (SES). SES includes factors such as income, education, type of neighbourhood, and whether there is a migration or refugee background. Hart and Risley (2003) reported what they called the *30 million word gap*, where they calculated that children of low SES heard approximately 30 million words less by age 3, compared to their high SES peers. The difference in lifestyle between low and high SES families caused this difference in linguistic input, and as a consequence, the lower SES children turned out to be less proficient in their L1 by the time they started school.

Hart and Risley (2003) conducted their study amongst American English speaking children within the USA. Immigration and especially refugee children can thus have a double setback in this respect; since refugees come into their new country with next to nothing, they often start out with a lower SES. It is difficult for their parents to find a job, even if they are highly educated, since in most cases they do not speak the language. Of course, this can improve over time. But for the children growing up in these situations, possibly having fled from a conflict zone and with parents struggling to fit in to a new culture, it is not unlikely that they already have some setback in their L1 development before they need to start learning Dutch at their new school. Finding the right methods to teach these children is crucial, as it is difficult to teach them in a standard classroom setting, where they cannot understand most of the instructions. And since we want to offer these children the best opportunities to learn and to advance in life, it is important to ensure they become linguistically proficient.

We need to find methods and factors that contribute to language learning, which are not based in a standard classroom setting, since classroom teaching is particularly challenging if you have insufficient knowledge of the language used at school. After all, you don't learn your first language in a school. Indeed, a reasonable place to start improving language learning for these children is to look at how children naturally acquire their first language, given that this occurs outside of school and it starts before children can speak or even understand the language. Even before Risley and Hart (2003) published their findings, a large number of studies had been conducted on this topic. In the last few decades we have learned a lot about which factors can facilitate language learning for children, in order to close the gap. These factors include: creating an immersive, motivating and fun learning

experience, movement, child-led activities, and outdoor environments. Since most studies attempt to find the optimal methods for language learning, they do not solely focus on teaching an L1, but can be applied to teaching an L2 as well.

Interestingly, most of the factors that were found to facilitate language learning, can also be mentioned as different aspects of one particular type of activity that children often engage in, regardless of their SES or cultural background: play. The idea of learning through play is not new, but finding types of play to use as methods for teaching is still relatively uncommon. This makes sense, since the factors found to facilitate learning are not exclusive to play (or language learning, for that matter), but they can be found as aspects of other activities as well. As such, they could be included in classroom teaching in different ways. However, one of the strengths of learning through play is the ability for play to combine all these factors together, to create an optimal learning environment.

Tomlinson and Masuhara (2009) conducted an extensive literature review on language acquisition through physical play. One of their main comments was that there were very few papers (up until 2009) that focused on this topic. They did find a number of papers that looked into separate aspects of play, such as the physical aspect of it, or that playing a game creates a new goal (winning) other than learning a language, providing a different kind of motivation. They found that there was a lack of literature that discussed research studies into using (physical) play in second language acquisition. To facilitate future studies, they provide a theoretical framework, supporting the use of games in L2 classes. The frameworks provided by Tomlinson and Masuhara (2009) consists of 6 “principles of language acquisition”, and a 9-step framework on how to implement games into the classroom<sup>1</sup>. The six principles highlight separate effects of gameplay on learning. When combined, they state that games provide an immersive experience in language use, that motivates the learners. Since the games require them to use the language in a different setting, the focus is shifted from feeling self-conscious about their language use to achieving a communicative goal. The learners are required to use language in order to win the game, which creates a deeper drive to understand the language, but in a fun way. To summarize, **one factor of learning through play that facilitates language learning is that it provides an immersive, motivating, fun, and meaningful experience of language in use.**

Another factor that can be a part of play and has been shown to advance learning, is moving. Singh and colleagues (2012) examined a large number of studies that explored this link between physical activity and academic achievement. Although they were tentative to make concrete claims, the body of evidence they investigated led them to conclude that there was a positive relationship between moving and learning. This claim is supported by a study conducted by Kirk and Kirk (2016), who compared a group of children that moved (marching in place, acting out verbs, jumping jacks) during normal academic classes to a group that just sat on their chairs during the same classes. They found significant improvements on a diverse set of literacy skills measures for the physical activity group compared to the sitting group. This provides evidence that **movement is another factor that facilitates language learning.**

Studies by Weisberg et al (2013), Hassinger-Das et al (2016), and Hopkins et al (2019) show that **child-led activities that are guided (but not controlled) by adults provide the most facilitating**

---

<sup>1</sup> The nine-step framework is not discussed here because it is catered towards adult L2 teachers and learners that are used to a classroom setting and share a common L1 in which they can discuss aspects of the lessons and specific grammatical structures they encounter.

**environment for learning.** Hassinger-Das et al (2016) mentioned that, in addition to being playful and motivating, games give the players some measure of control over their learning, which can increase their curiosity, tapping in to a deeper level of intrinsic motivation. Weisberg et al (2013) even went as far as to base their definition of play on being a child-led activity, putting the child in control. Doing so stimulates intrinsic motivation to participate, leading to deeper engagement in the activity. They mentioned that adult guidance during this play is a key component to facilitate learning. The child will have the intrinsic motivation to participate, but the adult is there to spot learning opportunities and to engage in these with the child, like providing labels for novel items or situations and spotting opportunities for symbolic thinking, for instance, a banana can function as a telephone. Hopkins et al (2019) agree that this *guided play* is the most beneficial form of play, but when children are unaccustomed to this type of play they might find it difficult to take the initiative. In these cases, *directed play* where the adult directs the activity to focus on learning opportunities, might be better.

In addition to these factors that are linked to the nature of play, another interesting view is to look at the environment where learning and playing takes place. A traditional learning environment is usually an indoor classroom setting. But when it comes to play, especially the types of play that could include a lot of movement, an outdoor setting like a playground is a feasible alternative. These **outdoor environments provide different types of interactions and experiences that can be very useful in learning.** Using outdoor environments in learning can make the lessons more immersive. This is supported by Acar (2014) who discusses the importance of outdoor spaces for the development of children, and Bussamante et al (2019) who developed outdoor language learning environments in low SES neighborhoods as a means to facilitate L1 learning. These environments, like a word game placed inside a bus stop for example, are meant to inspire meaningful interactions between children and their peers or caregivers, and thus facilitate learning.

What all of the discussed factors that facilitate language learning seem to have in common, apart from being related to play, is that they provide different ways to boost intrinsic motivation for learning, while providing a rich context in which to experience language in use, while at the same time simply being fun. Combining these factors into different interactive games designed for language learning, is precisely what was aimed for when creating the Noplica language playhouse<sup>2</sup>. It is a house-shaped structure, placed outside, that contains a number of interactive games that focus on different types or aspects of play and language. Originally it was created to stimulate the acquisition of English by children in rural India. After extended testing, at a Dutch elementary school and in the lab, it was found that creating a Dutch version in order to stimulate language learning for immigrant children in the Netherlands could be useful. For the current version, three games were translated from English into Dutch by Radboud University students.

All of the games incorporate the factors mentioned above, but to a different extent: The game *Energy center* focusses on moving while being immersed in language, as it requires children to use three hand bikes to listen to songs. The *Granny* game is more about interaction and immersion. It tells a story and asks the player to answer questions and repeat sentences, which are played back to the children. The *Dancefloor* game has a focus on vocabulary learning by linking images to spoken word, but it is set up in such a way that children are required to move around, interact with the game

---

<sup>2</sup> The Noplica language playhouse is designed by Entwerpen design agency, on behalf of the ChildTuition foundation, in collaboration with the Radboud University and the Baby and Child Research Center. The version used in the current study has been built and installed by Houtplezier at OBS de Bloemberg in Nijmegen.



by pressing the correct buttons, and to work together to win. Both the Granny game and the Dancefloor game provide positive feedback, similar to the kind of guidance an adult might provide during play. All of the games are in an outdoor, (language) immersive and fun environment. The children can decide which games to play and how, making the Noplica playhouse a child-led learning experience, with the set structure in the play substituting adult (or teacher) guidance.

All of the mentioned factors have shown to contribute to success, and are present in the playhouse, but the question still remains if this approach actually works: do the children enjoy the games, and, most importantly: do they actually learn from them? The current study aims to answer this question for the Dancefloor game. This game was selected because it combines all of the mentioned factors that facilitate learning, and it focuses on vocabulary learning, which is a relatively easy aspect of language to measure.

The hypothesis is that the Dancefloor game will succeed at increasing both perceptive and productive vocabulary knowledge, due to the nature of the game; it immerses children in the linguistic environment in a way that brings them joy and thus motivates and stimulates, while they are moving, cooperating and receiving positive guiding feedback as part of the play. The game will be described in detail in chapter 2.2.1.

In order to test the effectiveness of the Dancefloor game we need to assess the participants vocabulary knowledge. This requires creating a vocabulary test that uses the words from the Dancefloor game. One group of children is selected to play the Dancefloor game, and compared to a control group playing a different game. This will single out the effects of the Dancefloor games while keeping all other factors relating to the study as similar as possible. It is also necessary to find out how the children respond to playing in the Noplica language playhouse, and if this effects their normal classes. Hence, the present study is a small-scale pilot study, aiming to develop the necessary materials and methods to ultimately answer the main question in a larger study.

The remainder of this thesis will describe the development of the vocabulary test as well as the experimental methods and procedures in Chapter 2. Chapter 3 will focus on the results of the experiment. Chapter 4 will discuss these findings in light of the literature, provide suggestions for improvements to the used procedures and materials, and mention some aspects that can be further explored in future studies. Chapter 5 will conclude the findings in this paper.

## 2 Method

### 2.1 Participants

There were twelve participants. Seven were female. Age ranged from five years and three months to seven years and five months. The children have lived in the Netherlands for a relatively short time, varying from four months and one day up to 14 months and 29 days, with a mean of ten months and 21 days. The children's age ranged from five years, three months and nine days to seven years, five months and 29 days, with a mean age of six years and 16 days. The children came from various linguistic backgrounds, speaking different L1's.

Participants were picked from three different classes at OBS de Bloemberg, Nijmegen. This school specialises in teaching Dutch as an L2 to immigrant children, while following a normal Dutch curriculum. Four children were in one second grade class (2a), four children were from another second grade class (2b) and four children were in the third grade. Six children were assigned to the test group; two from each class. The remaining six were assigned to the control group. An overview of the participant information can be found in Table 1.

Note: because one participant (number two) was not present for the post-test, this data was excluded from the final analysis. However, since this participant was present for the pre-test and most of the play sessions, his data is included in the overview as well as any figures and discussion of the observation data.

**Table 1: Participant information overview**

PP No.	Sex	Mother tongue	Grade	Group
1	f	Tigrinya (Eritrea)	2	test
2	m	Arabic (Syria)	2	test
3	f	German	2	control
4	f	Turkish	2	control
5	m	Arabic (Morocco)	2	test
6	f	Turkish	2	test
7	m	Arabic (Syria)	2	control
8	f	Vietnamese	2	control
9	f	Turkish	3	test
10	m	Mandarin (China)	3	test
11	f	Mandarin (China)	3	control
12	m	Arabic (Syria)	3	control

### 2.2 Materials

#### 2.2.1 The Noplica language playhouse

##### 2.2.1.1 Test group game: Dancefloor

The Dancefloor game consists of two walls, placed in a 90-degree angle. On each wall are a computer screen, three touch buttons with a hand-shape on them and blue lights, and two speakers. The game consists of two consecutive rounds, and round one starts when one of the buttons is touched. Both screens will show a different image, from the same theme (out of 20 themes available for the game). A short sentence is played, containing the name of one of these images. For example, an image of a chair and an image of a table are shown on the screens, and the audio will play: "Waar is de tafel?"

(*Where is the table?*). At that point, all of the buttons will start to flash their blue lights and the goal is to push a button on the 'correct' wall, i.e., the wall with the image of the table, within the time limit. When a correct button is hit, the game plays a positive feedback sentence, for example: "Goed gedaan! De tafel!" (*Well done! The table!*). There is no response to pressing incorrect buttons. When a theme is completed (10 correct responses on a row; there are 10 items per theme, and each one is asked once), a short song will play and all the words are repeated, while a short cartoon clip showing all the items is played on the screen. After this, round two begins. In this round, both screens will show the same image, and the sentences played are similar<sup>3</sup> to the first round. But for this round, only one of the buttons will flash blue lights, and the goal is to find and hit this button within the time limit. Again, it does not matter if incorrect tiles are also pressed. All 10 items from the current theme appear again, in random order. When all items are played correctly in round 2, another song and cartoon clip will play, repeating all the words again. After that, a new round is started, with a different theme. There are 20 themes in total, and they are played in random order.

For the first game/theme, the response time limit is 12 seconds, and it will get half a second shorter after each time both the first and second round of the game are finished completely. When the correct tile is not pressed on time, it is game over. The game resets the timer and goes into "pause" mode, where the Noplica logo is displayed on both screens. A new game starts again when one of the buttons is touched. A complete list of all themes, items and sentences can be found in Appendix A: Dancefloor content.



Figure 1: The Noplica language playhouse, with the Dancefloor game (in orange) and Energy Center (in blue).

<sup>3</sup> The actual sentences are the same, but their order is random, so a different sentence can occur with the same item; where in round 1 you would hear "Waar is de tafel?" (*Where is the table?*), in round 2 you could hear "Kun je de tafel vinden?" (*Can you find the table?*)

### 2.2.1.2 Control group game: *Granny*

The *Granny* game consists of one wall with a computer screen, a large wooden wheel, two speakers and a microphone. The game is made up of different short stories or scenarios, concerning the two main characters: *Granny* and *Birdy*. Within each story, children will have to answer some questions (forced choice), by turning the big wheel in the direction of the picture depicting the correct answer. They are also asked to repeat (parts of) some sentences, and their responses are recorded and played back to them, both immediately and at the end of the story. There are 20 stories in total, and the amount of questions, repeated sentences and the type of feedback varies per story. The stories are played in random order. The game begins when the wheel is turned. When any question is answered incorrectly, this is explained, and the question is asked again until a correct response is given. For example, the game might ask “*Waar is de berg met de stenen?*” (*Where is the mountain with the rocks?*), while showing a mountain with rocks on the left and a mountain with snow on the right. If the wheel is turned to the right, the game would answer: “*Nee, dat is niet de berg met de stenen. Dat is een berg met sneeuw. Probeer het nog eens, kun je de berg met de stenen vinden?*” (*No, that’s not the mountain with the rocks, that’s a mountain with snow. Try again, can you find the mountain with the rocks?*). When the children are asked to repeat (parts of) a sentence, they always receive positive feedback, even if they haven’t said anything at all. Sometimes, they are asked to repeat it again, but this is set for each scenario; the feedback is not responsive to the actual input of the children. For example, *Birdy* might say something like: “*er staan vijf kaarsjes op de taart. Kan jij dat zeggen? Een taart met vijf kaarsjes*” (*there are five candles on the cake. Can you say that? A cake with five candles*). The children would then most likely attempt to repeat “*een taart met vijf kaarsjes*”, but they could also be quiet, or scream, or make whatever sounds they want. The game could then ask: “*kun je dat nog eens zeggen?*” (*can you say that again?*), or it could move on to feedback: “*goed gedaan! Een taart met vijf kaarsjes!*” (*well done! A cake with five candles!*). Sometimes, the game then plays the audio recording of what was said, and for other stories, it waits until the end of the story and then plays back all the recordings, stating, for example: “*goed gedaan! Dit zeg je wanneer je een schat zoekt:*” (*well done! This is what you say when going treasure hunting:*) [recording is played].

Children can continue playing the stories for as long as they like; there is no game over. The game resets after a while if the wheel is not turned when required. It goes into a “pause” mode where the *Noplica* logo is displayed on the screen, as can be seen in figure 2, and starts a new story when the wheel is turned again.

The main characters in all the stories are *Birdy*, a blue bird that is also present in the *Noplica* logo, and “*Oma*” (*Granny*). *Birdy* lives with *Granny*. Most of the scenarios are common situations or actions that occur in daily life, like getting dressed for hot or cold weather, taking a trip to the zoo or meeting new neighbours. But they also fit in well with children’s stories and imagination, like going on a treasure hunt, or having an underwater adventure. An overview and brief summary of all 20 stories can be found in Appendix B: *Granny* stories.



Figure 2: The Noplica language playhouse, showing the Granny game (in green). This game is on the opposite side of the Dancefloor game.

## 2.2.2 Vocabulary tests

### 2.2.2.1 Vocabulary perception test

The materials developed and used for the vocabulary perception test were a flip-book of 20 pages, and a set of scoring sheets.

The flip-book contained 20 laminated A4 pages. On each page were four images, consisting of two target item-distractor pairs from two different themes that are present in the dancefloor game<sup>4</sup>. The experimenter asked the children to point to the picture of the word that was mentioned. After the child pointed to one of the pictures, the experimenter turned the page and the procedure was repeated. The flip-book was run through twice, but in the second round a different picture was named. An overview of the 20 pages can be found in Appendix C: Vocabulary perception test.

The response (correct/incorrect) was noted on a scoring sheet. There were two versions, alternating in the order in which the target items are asked. Both versions were evenly distributed over the participants. Participants that got version 1 in the pre-test, got version 2 in the post-test (and vice-versa). Both versions of the scoring sheet can be found in Appendix D: Scoring sheets.

The 'raw' score for the vocabulary perception test is the number of correct responses. For analysis, the percentage of correct answers was calculated for the perception vocabulary test.

<sup>4</sup> A complete list of all 20 themes and 201 words is included in Appendix A: Dancefloor content.



### 2.2.2.2 Vocabulary production test

The materials for the vocabulary production test consisted of an A3 laminated paper with 16 images, and 16 smaller cards containing those same images. The same scoring sheets were used for both the perception and production vocabulary test. A smaller image of the original A3 page can be found in Appendix E: Vocabulary production test.

The A3 paper was placed atop a table and the participant was asked to lay out the corresponding smaller cards on top, naming each item as the card was placed. Correctly named items were noted on the score sheet (correct/incorrect). Order of items was determined by the participant (the deck was shuffled but the participant could look through it freely, or lay them out on the table). The session was recorded to check responses in case of doubt, and to keep the option to look at individual linguistic differences (pronunciation, word choice etc.), although this was not done for the present study.

The raw score for the vocabulary production test was the number of correct productions. An answer was scored as correct, when the participant used a common Dutch word to name the object, and that word described the target object. In case of doubt the researcher prompted the participant to repeat or possibly specify their answer. For example: if “vinger” (*finger*) was the target word, but the participant said “hand” (*hand*), the researcher agreed that it was a hand, but pointed to the finger and asked if they also knew what that was called: “Ja, dit is een hand, maar weet je ook hoe dit heet?”.

Differences in pronunciation were disregarded as long as the word was very close to the Dutch pronunciation. To give an example: if the target word was “T.V.” /te:ve:/, but the participant said “televisie” /te:ləvizi/, this was counted as correct. If they pronounced it slightly different, for example with an –on /te:ləvɪzən/ ending, it was still counted as correct. If more than one sound or part of the word was pronounced differently, for example an entirely English pronunciation /tələvɪzən/, it was not counted as correct. If there was any doubt, the researcher noted this on the scoring sheet and listened to the recordings at a later time, to determine whether the answer was correct or incorrect. Deviations from the Dutch pronunciation were not very frequent; usually it was quite clear whether the participant attempted to say the correct word or not. But for words like “televisie” or “chocolade”, the word in the child’s native language can be quite close, or they might know it in English, hence the focus on them attempting to pronounce the word in a close-to-Dutch manner. When it was clear that the pronunciation was an attempt to say a Dutch word rather than a related word in a different language, it was counted as correct.

### 2.2.2.3 Test material selection procedures

The images used in the vocabulary tests were selected from the set of images used in the first round of the dancefloor game. There is a total of 20 themes containing ten words each<sup>5</sup>. In order to select the appropriate words for the target and distractor items, the corresponding images were checked for recognisability. Teacher questionnaires were used to determine whether the participants would already understand and/or use these words, and whether they would be included in the regular school curriculum during the time of the study. An attempt was made to differentiate between “easier” and more “difficult” words, by checking them against the PPVT-NL-III (Dunn & Dunn (2005).

---

<sup>5</sup> There are 11 words within one specific theme, where “teacher” from the English version of the game was translated as both “juf” (female teacher) and “meester” (male teacher) for the Dutch version, making 201 words total.

Finally, item-distractor pairs were carefully selected and distributed. All of these steps are further explained below.

### Image recognisability

To control for image recognisability, i.e. whether the images actually portray what they are supposed to in a way that is clear to children of the target age (4-6 years), all images were shown to a native Dutch speaker of a comparable age and education level (F, age 5;11, second grade primary school) in random order. She was asked to name the pictures. Her answers were noted down and later compared to the target words. Pictures that were difficult to recognise were excluded from the set.

Some cultural differences between the originally intended target group of the Noplica house (children in rural India) could in part be the cause for this lack of recognisability. For example, a typical Indian farmer, as seen on the image in the Dancefloor game, does not match the typical Dutch image of a farmer, who would usually be depicted wearing clogs and an overall. And the images of a primary school, restaurant or shopping center are also very different from typical Dutch examples of these places. Notes on these differences will be shared with the developers of the Noplica playhouse, so that these images can be updated to better suit the Dutch version of the game.

### Teacher questionnaire

To find out whether the participants would already be familiar with some of the words in the game, a questionnaire was provided to the teachers of the participant groups. The questionnaire consisted of all 201 words from the game, and teachers were asked whether most of the students would *know* (receptive knowledge), or *know and also be able to say* these words (receptive and productive knowledge), prior to the start of the study.

The teachers were also asked whether any of the words would be used in the normal curriculum during the research period, or whether they might fit in with a “theme” that the classes were working on. It is not uncommon for Dutch schools to arrange lessons around a theme, such as “being ill”, where they might learn words for body parts, how to express things like headache or stomach ache, what a visit to the doctor or dentist might look like, what an ambulance is, etcetera. Classes will conduct different activities, at their own level, in association with the theme, like arts and crafts or playing with suitable toys. During such a theme, the normal curriculum (lessons in language, math, etcetera) is also followed, and this does not necessarily match with the current theme, which is why the questionnaire asked about both the normal curriculum as well as any themes.

Prior to the start of the current study, the original plan was to have all the children in one particular class (the 3<sup>rd</sup> grade class) participate. Their teacher filled out the questionnaire before the vocabulary tests were developed, and her answers were thus used to classify the items from the Dancefloor. The words that most children already *know* or *know and say* were counted as “easier” words for the development of the test, and the remainder was counted as more “difficult”.

However, due to the COVID-19 pandemic, in March of 2020 all schools were closed and the study was put on hold. At the end of May it became clear that the study could start up again, but not all children were present at the school, so the group of participants changed to include the younger children, in order to have enough participants. The teachers for all three participating classes filled out the questionnaires again during the study, so that their answers might be used to control for words the participants might have picked up from the regular curriculum or from working within a

specific theme, rather than learning it from the playhouse. Due to the pilot-test nature of the current study, the large individual differences and the small participant group, this data was not used in the current analysis, but it might be useful in order to control for any differences between classes in future studies.

### Peabody Picture Vocabulary Test

Not all words from the Dancefloor game might be equally easy to learn. This can depend on many different factors, such as which L1 a participant has, their interest in any particular subject or theme, some factors about the concept or item the word refers to, or some factors about the word form itself that make it more difficult to learn. This is a complicated set of factors, making it difficult to judge which words might count as easier or more difficult. However, some kind of measure of this was required, to ensure that the vocabulary tests would both include some words that the children would most likely already know, as well as some words that they were very likely to learn and some words that they are less likely to learn. This balance is necessary, since it is very demotivating for a child to take a test and not know any of the answers, so you do not want it to be too difficult during the pre-test. At the same time the test is intended to show how much they can improve their word knowledge, so there need to be enough words that they can still learn.

The Peabody Picture Vocabulary Test (PPVT) was developed as a measure to assess receptive vocabulary, normed by age. The currently used version, the PPVT-NL-III (Dunn & Dunn (2005), is suitable from age 2;3 and up to over 90 years of age. For each age-group, there is a so-called entry-level for the test<sup>6</sup>, and all words before that entry level are assumed as known to that specific age group. When an individual makes more mistakes than expected for their age group, they move down a level until they give enough correct responses. As such, for children, the PPVT can be used to measure a delay in word learning, but it is also very clear that the words used in this test are of increasing levels of difficulty, and this has been tested and adjusted over many versions of the PPVT, meaning that the current levels are quite robust, and can thus be used to check whether a specific word might be easier or more difficult to acquire.

All the words in the Dancefloor game were compared against the PPVT-NL-III. If they were included in the PPVT, it was noted in which subset they were found, which tells us at which age they are most commonly acquired by Dutch L1 children (without any language learning difficulties). The words up to set three (entry level of the PPVT for children up to age 3;11) were classified as “easy”, and the rest was classified as more “difficult”. This cut-off point was chosen because it is in line with a pre-elementary school age (and corresponding average vocabulary level) for Dutch children, which matches the intended level of the playhouse.

Only 61 out of the 201 words from the Dancefloor game were included in the PPVT, so we relied on the teacher questionnaire alongside the PPVT levels to determine whether a word could be classified as “easy” or more “difficult” to learn. In the tables in Appendix F: Test item selection brainstorm, it is noted which words were found in the PPVT, and in which subset. It is also made clear in this appendix which words were classified as “easy” and which were classified as “difficult”.

---

<sup>6</sup> There are two practice sets, and then 14 sets of increasing difficulty. The first set is the entry level for age 2;3-2;5. The 14<sup>th</sup> set is the entry level for age 36 and above.



### Item-Distractor pair selection

It is important to be careful when selecting pairs of target items and suitable distractor items. The distractor needs to seem like a plausible candidate, but it should not make it too difficult to recognise the intended target. Most commonly, distractor items are chosen from a similar semantic category, and can be paired on phonetic likeness as well. However, these characteristics only refer to the word pair, and they do not look at their corresponding images. Delle Luche et al (2015), although they were investigating the best methods for using the Intermodal Preferential Looking paradigm, looked into the quality of visual stimuli. They mentioned how different studies use 'visual salience' as a selection criterion. The concept of salience is not a very clear one. One explanation is that something that is highly salient draws more attention compared to something that is low in salience. This is a difficult thing to measure, since it is often based on more subjective measures, or even personal preference. For example, a fire truck and an ambulance can both be of similar size, have a bright colour, and flashing blue lights. Thus, one might argue that they are similarly visually salient. However, if your favourite colour is red, you might be more drawn to the fire truck. Delle Luche et al (2015) attempted to use an objective measure for visual saliency, by calculating the luminance of each pixel in an image into a vector, and then performing cross-correlations of image vector pairs. This is complex to measure and calculate, and this was not used for the present study. The present study does attempt to take visual saliency into account, by comparing possible distractor images to the target item, based on size, shape and colour. For example, both the images for cookie and for potato are of a very similar colour, size, and shape, thus making an excellent item-distractor pair based on image saliency, even if their word forms are dissimilar and they also are not that close based on semantic relations; although both are a food, one is sweet and a snack and the other one is more savoury and eaten at dinner.

To select item-distractor pairs for the vocabulary perception test, all items (except those that were excluded based on the image recognisability check) from each theme were compared. First, it was assessed if any two words were phonetically close, such as "kop" and "kom" (*cup* and *bowl*). Secondly, it was assessed if any two items were closely linked semantically, such as "potlood" and "pen" (*pencil* and *pen*), since they can both be used to write with. Thirdly, all images from the same theme were compared, to see if any two had a close likeness, such as "koek" and "aardappel" (*cookie* and *potato*) mentioned above, but also "paard" and "ezel" (*horse* and *donkey*), which are both semantically and visually related. Using these three criteria, a total of 40 item-distractor pairs was selected, two from each available theme in the Dancefloor game.

For the vocabulary production test, a selection of 16 target items was made. This number was chosen so that the images would fit in an appropriate size on an A3 sized page. There are 20 themes in the Dancefloor game. One theme, "places to go", was excluded. The images from this theme do not show a single item, but are a picture of a place, like a farm or school, placed in its surroundings. The inclusion of background in these items is necessary to make them recognisable, but it causes them to stand out too much from the pictures in all other themes; it provides a very different level of saliency. Additionally, the image recognisability test showed that most images from this theme were very hard to recognise. This leaves 19 remaining themes, to select 16 items from. There are four themes that focus on food items, and four themes that focus on animals. To ensure all semantic categories were present in the vocabulary production test, the logical step was to exclude some of these overlapping themes. In order to maintain the balance in easy and difficult items, three animal

themes were selected (since these provided a higher number of easy items), and two food themes. One target item was selected from each of the chosen themes, providing 16 images.

### Item distribution

After the pairs of target and distractor items for the vocabulary perception test were selected, these pairs were checked for an even distribution between the “difficult” and “easy” words. This was done in such a way that there was a relatively even distribution of easy-easy (6 pairs), easy-difficult (9 pairs), difficult-easy (9 pairs) and difficult-difficult (16 pairs) item-distractor pairs. The larger amount of difficult-difficult pairs is caused by a lack of “easy” items within certain themes, mainly the food categories. An overview of the selected items and their distribution in hard-easy categories can be found in the table on page 42, in Appendix F: Test item selection and brainstorm.

It was considered to distribute the items over the vocabulary perception test in such a way that it would increase in difficulty, much like the PPVT. However, it was unclear whether items labelled as “easy” were indeed easier or more well known by all participants. Additionally, it was important to make sure the children stayed motivated while taking the test. Making the test increasingly difficult, when it could have been quite difficult for some participants to start with, could have been a demotivating factor. Thus, the difficulty levels were evenly distributed. Item-distractor pairs were also distributed over the 20 pages in such a way that there was no semantic overlap between the two sets that would appear together on each page in the test; food items were never paired with other food items, animals not with other animals, etcetera. This ensured that, even though all target items appeared on a page with three other images, only one was a likely distractor, and two of them were more unlikely distractors, providing a relatively even level of difficulty for each page.

The vocabulary production test consisted of 16 items. Out of these 16 items, 8 were classified as “difficult”, and 8 were classified as “easy”. All items for the production test were distributed over the A3 page semi-randomly, so that the easy and difficult items were evenly distributed, but not in any particular order. An overview of this distribution can be found in Appendix E: Vocabulary production test.

## 2.3 Design

The **dependent variables** used in analysis were calculated from the difference between the post-test scores and the pre-test scores for the vocabulary perception and production tests. This score was named *improvement*. The percentage correct for the perception (percentage of correct answers out of 40 items) and production (percentage of correct answers out of 16 items) tests were calculated, for both the pre-tests and the post-tests. The pre-test scores were then subtracted from the post-test scores, giving the *production test improvement* and *perception test improvement* scores. A *mean improvement score* was also calculated, to look at overall improvement. This was done by adding the perception and production percentage scores, and dividing this by two. This method was chosen so that both parts of the test would have equal weight, since there are much fewer items for the production test compared to the perception test. Percentage was used rather than raw scores to make the scores more comparable for future research.

The **independent variable** used in analysis is the *group* variable, consisting of the *test group* that played the Dancefloor game, versus the *control group* that played the Granny game.

For the intended future study, using a larger number of participants, additional factors should be considered. These factors include participants' *language background*, their *length of stay* in the Netherlands, and their levels of *enthusiasm* and *focus* during the play sessions. For the current study, these factors were not included because of the low number of participants. Adding too many factors to such an analysis can lead to loss of power, and can make the results harder to interpret when there are large individual differences. These individual differences tend to even out over larger groups, yielding clearer effects of such factors. But these effects can be impossible to detect over smaller sample sizes.

**Participants were evenly distributed** over two groups, the test group and the control group. This was done for each class. There was a total of 12 participants, four children from each class, two of which were in the test group and two were in the control group. Where possible, gender was evenly distributed. There were many different language backgrounds, and children were paired in such a way that they did not share the same native language. This ensured that the children would communicate in Dutch during the play sessions. There were also large individual differences in the children's length of stay in the Netherlands, ranging from four months and one day up to 22 months and one day. For the current study, this was not taken into account when distributing the participants over the groups. However, for future research it is recommended to do so, since it appeared from the data that a shorter stay in the Netherlands might correlate to larger improvement scores. Details on the individual participants language background, their length of stay in the Netherlands and which group they were assigned to can be found in the participant information overview in Table 1 (p. 6).

## 2.4 Procedures

Before the start of the study, the school granted permission to conduct the study. The director, Iris Kokosky Deforchaux, signed a consent form. This granted permission for the researcher to administer the vocabulary tests and conduct the play sessions with the pupils. Details on dates and times for the test- and play sessions were discussed with the relevant teachers. The parents received implicit consent forms, because their levels of Dutch and English would not always suffice to understand an informed consent form. In summary, the parental consent forms stated that there would be a study, that their child would participate anonymously, and what they could do if they did not agree to this. Both the signed school consent form and the parental consent form can be found in Appendix G: Consent forms.

Upon the start of the study, all participants completed both the vocabulary perception test and the vocabulary production test. Their scores were noted on the score sheet, and the session was recorded (audio only). Afterwards, there were eight play-sessions, twice a week for four weeks, in which the participants played their assigned game for 15 minutes, while the researcher observed these sessions. Finally, the vocabulary perception and production tests were administered a second time. Again, scores were noted on a score sheet, and the session was recorded (audio only). The improvement scores for the analysis were then calculated from the percentages correct on the vocabulary tests.

### 2.4.1 Vocabulary pre-tests

All participating children were individually taken from the classroom by the researcher and brought into a separate room, with a table, two chairs (next to each other) and audio recording equipment.

The task for the vocabulary perception test was explained to the child, in such a way that the researcher asked for the child's help in making sure some images were clear enough. The researcher asked the child to point out something: "*Kun je voor mij de/het ... aanwijzen?*" (*can you point to the ...*). Two practice pages were presented, after which the child could ask any questions. Then the actual task began. All 40 target items were asked, and the researcher noted the responses on the score sheet. No comments on correctness were made, the child was thanked for their answer.

The second part was the vocabulary production test. The researcher put the A3 page on the table, and handed the child the set of cards (in random order; they were shuffled). The researcher asked the child to look through the images, and see if they recognised any. The child was then asked to name this item in Dutch, and place it in its correct spot on the A3 page. The researcher noted their responses on the score sheet. If the child did not know the Dutch word, the researcher asked them what it is called in their native language, and tried to repeat this. This was done to make the test more interactive and game-like. When all 16 cards were in their correct places, the test was complete.

During both tests, the researcher used a number of tactics to interact with the children, to keep them engaged and motivated. She used gestures, made some animal noises, asked participants if they also have a tongue (and stuck it out), showed them funny socks she was wearing, etcetera. During the vocabulary production test, the researcher also asked some additional questions to prompt a more accurate response, when necessary. The researcher was careful with these interactions, making sure they would not influence the responses, but rather used them as reaction to the responses. "Yes, thank you for pointing at the socks, do you also have socks on? Look, mine have a pig on them." "Ooh, that is a nice animal! Do you know what it's called? No? what sound does it make? Does it go like this? (Makes elephant noise while using one arm as a trunk) Can you do that?"

#### 2.4.2 Play sessions

The play sessions took place twice a week for four weeks in a row, on the same days: Tuesdays and Fridays for the second grade groups, and Wednesdays and Fridays for the third grade. The participating children were divided into groups of two, so that they were able to play the playhouse games together. The researcher collected them from their class, and brought them out to the playhouse. If one student from a particular group was absent, the remaining student was allowed to pick a non-participating classmate to pair up with for the play session, so that they would not have to play alone.

The children were instructed to play their respective game, and the researcher showed them how to press the buttons or turn the wheel, but they received no further instructions on how to play the games. The researcher observed the session, only intervening if the children stopped playing the game or in case of any interruptions (toilet breaks or emergencies). A number of observations, such as the level of enthusiasm and focus of the participants, were noted on the observation checklist, which can be found in Appendix H. After 15 minutes of playing, the researcher asked the children to stop, and brought them back to their class.

Initially, the plan was to have the control group play the Energy Center game, where they use hand bikes to turn on lights and play songs. However, during the first few play sessions it became clear that it was too tiring to do this for the full 15 minutes, and the children would not play this game for more than a couple minutes before running short of breath and attempting to go play elsewhere.

Thus, the researcher opted to let them play the Granny game instead. This turned out to be a good decision since most participants really enjoyed this game and played it easily for 15 minutes, which was very comparable to how the test group children had responded to the Dancefloor game. However, there are words that appear in both the Granny game and the Dancefloor game, and in the current study, this was not taken into account when selecting items for the vocabulary test. As a result, 21 words used as items in this test also occur in the Granny game. This could weaken any found results, but removing these items from analysis would result in deleting about 35% of the data for the vocabulary tests. Since there is already a limited amount of data due to a small number of participants, it was decided to include the overlapping words in the analysis of the present study. However, for future research, it is advised to adapt the vocabulary tests in order to exclude these words.

### 2.4.3 Vocabulary post-tests

The participants took the vocabulary tests again. The procedure was the same, but the vocabulary perception test versions were altered, meaning that the target item order was different. This, in combination with the lack of comments on correctness during the first test session, should prevent any learning effects from the tests. This time, during the test there was still no feedback on correctness, but afterwards the researcher asked the children if they wanted to know the words for the items they could not name. They received a colouring page with characters from the playhouse (Birdy and Oma/Granny) at the very end, to thank them for participating. Images of these pages can be found in Appendix I: Colouring pages.

## 2.5 Apparatus

**Audio recording equipment:** The researchers' mobile phone: a Samsung Galaxy S7. The standard voice recorder app was used on the "interview" setting. Files were stored in m4a format (MP4 for audio only).

**Test scoring sheet** for the vocabulary tests, in two versions, as can be found in Appendix D.

A **Play session observation checklist** was used to make consistent notes during the play sessions. It noted the current session date, time and participants and the start- and ending time for the session. On a seven-point-scale, the researcher noted how *enthusiastic* and *focussed* the children were, to what extent they were repeating words out loud, and to what extent they involved the researcher in their game. There was also room to note if they played the game as was intended, and if anything else stood out, either for the group or for individual children. The full observation checklist can be found in Appendix H.

A **Classroom questionnaire** was used to assess if and how the play sessions were affecting the participants' behaviour in the classroom. The respective teachers filled one out for each student, and noted if their behaviour was different on a play session day compared to a normal day. On a five-point-scale, they could note any differences with regards to how quiet and how focussed the children were. There was also the question of whether the child seemed to like the study, and how this became apparent, and there was room for any other comments. A short version of the classroom questionnaire (in Dutch) can be found in Appendix J.

## 2.6 Analysis

To see if the improvement scores for the test group were significantly different from the improvement scores of the control group, a MANOVA (multiple analysis of variance) was used, with the improvement scores for the production and perception test and the mean improvement as dependent variables and group as the independent variable. First, descriptive statistics were explored to see if there was any difference between the groups. Then, the data were checked to see if they met the assumptions for running a parametric test.

The descriptive data from the play session observation checklists was also explored. Although no analysis was currently conducted with this data, it is still interesting to see if there are any indications that these variables could be of influence and are thus useful to add to in to the analysis of any future study.

### 3 Results

In paragraph 3.1 the play session observation checklists and the classroom questionnaire findings will be reported. These data were not added to the statistical analysis, due to the small number of participants. The vocabulary test scores will also be reported in this paragraph. Paragraph 3.2 explains how these data were checked for meeting the assumptions for parametric testing, and paragraph 3.3 contains the results of the analysis: a MANOVA with follow-up Univariate ANOVA's.

#### 3.1 Descriptive statistics

##### 3.1.1 The play session observation checklists

The observation checklists filled out during the play sessions contain data on how enthusiastic and focussed the participants were, to which degree they repeated words out loud, and if they involved the researcher in the session. All these were scored on a scale of 1 to 7 for each participant individually, during all play sessions. A score of 1 means no enthusiasm/focus/repeating/researcher involvement, and a score of 7 means very enthusiastic/focussed, constant repeating/involving the researcher.

For focus and enthusiasm, the researcher based the scores on how the individual child's behaviour compared to the whole group. A score of 4 represents average focus and enthusiasm.

The measure for repeating words was scored slightly differently per game, as the games had different requirements. The Granny game actively asks the players to repeat (parts of) phrases. When participants did only this, it was scored as 4 (average). The Dancefloor game does not require any repeating, but the songs played between the different stages of the game do invite to repeat the words here, so if the children repeated words only during these parts, this was scored as 4 (average).

The researcher involvement scores represent a more absolute level of how much the children would involve the researcher in their activity. A score of 1 would indicate that absolutely no involvement was required, and a score of 7 would indicate that the children had asked the researcher to participate in the game (though neither occurred.) In the case of minor interventions or some encouragement, a score of 2 to 4 was given. There were occasions where the children would start a conversation with the researcher. This was scored as 5 or 6. The researcher attempted to minimise all involvement, without being unfriendly or disrespectful towards the children.

Figure 3 visualizes the mean observation scores for each observed factor per group, over all play sessions and participants. The error bars show 1 SD, to indicate individual variations. As can be seen from this figure, this variation was relatively small for focus (overall range: 2.0) and enthusiasm (overall range: 1.5), slightly larger for repeated words (overall range: 3.0), and quite large for researcher involvement (overall range: 3.8).

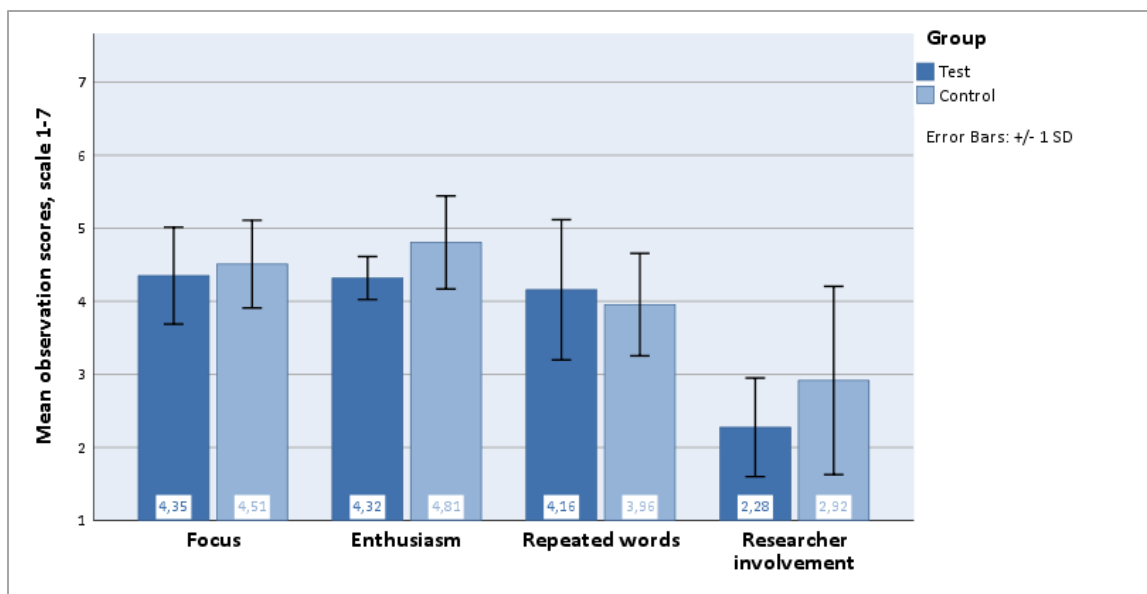


Figure 3: Summary of the observation checklist data, group means over all play sessions, per observation item.

### 3.1.2 The classroom questionnaires

Overall, the classroom questionnaires showed no influence of the play sessions or the presence of the researcher on the classroom behaviour of the participants. Two pupils became slightly more excited when the researcher was present, but this did not influence their in-class performance.

### 3.1.3 The vocabulary test scores

There was a large spread in individual test scores. For the pre-test, the lowest overall score was 18.8% correct, and the highest score was 85.6% correct. For the post-test, the lowest overall score was 48.8% correct, and the highest score was 88.8% correct. Although it was expected that the post-test scores would be higher than the pre-test scores, one participant actually scored 5% *lower* during the post-test compared to the pre-test. This can also be seen in Figure 4, where the pre- and post-test scores for all participants can be compared. The mean and SD for the test scores per group and in total, over both production, perception and mean (overall) vocabulary tests can be found in Table 2.

Since the goal was to see how much the children learnt during the play sessions, their improvement was calculated, by subtracting the pre-test scores from the post-test scores. These improvement scores for the vocabulary perception test, vocabulary production test, and overall improvement can be compared in Figure 5. This graph shows the mean improvement scores per group. The 1 SD bars indicate the large amount of individual variation. As can be seen in Figure 5, the mean improvement scores for the test group were higher than those of the control group. Statistical analysis is required to show if this difference was significant.



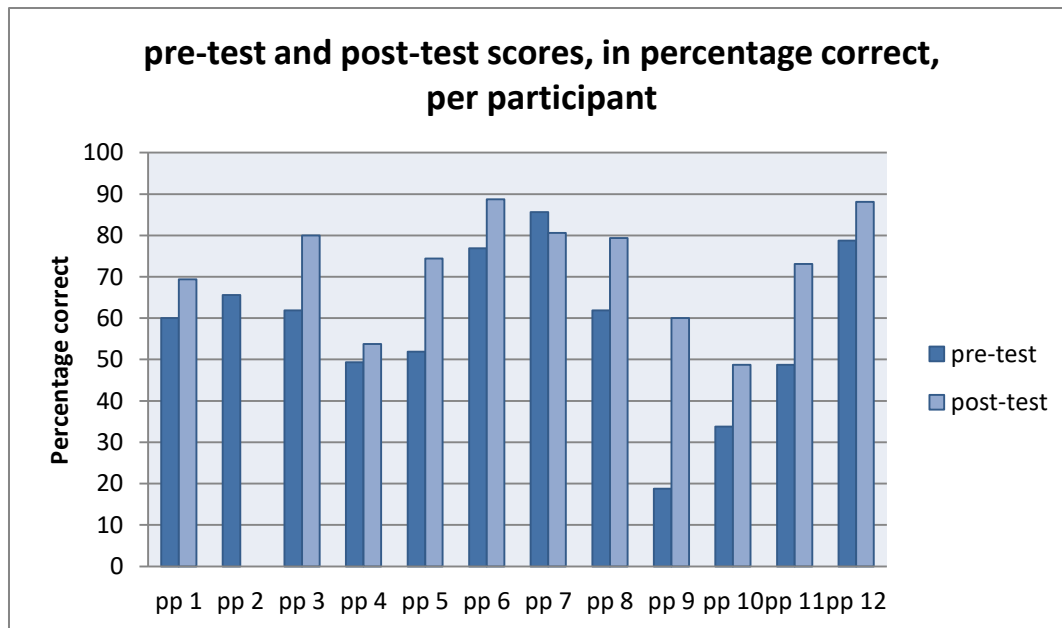


Figure 5: Bar chart indicating the individual pre-test and post-test vocabulary scores (perception and production combined).

Table 2: The mean score (and SD) for all vocabulary tests, per group and over all participants.

Vocabulary test	Test group score	Control group score	Mean score
Perception pre-test	59.6 (22.3)	72.5 (11.3)	66.0 (18.1)
Perception post-test	76.5 (11.7)	85.0 (9.7)	81.1 (10.9)
Production pre-test	42.7 (20.7)	56.3 (19.8)	49.5 (20.5)
Production post-test	60.0 (19.1)	66.7 (15.1)	63.6 (16.5)
Overall pre-test	51.1 (21.5)	64.4 (15.1)	52.3 (23.9)
Overall post-test	68.3 (15.1)	75.8 (11.8)	72.4 (13.3)

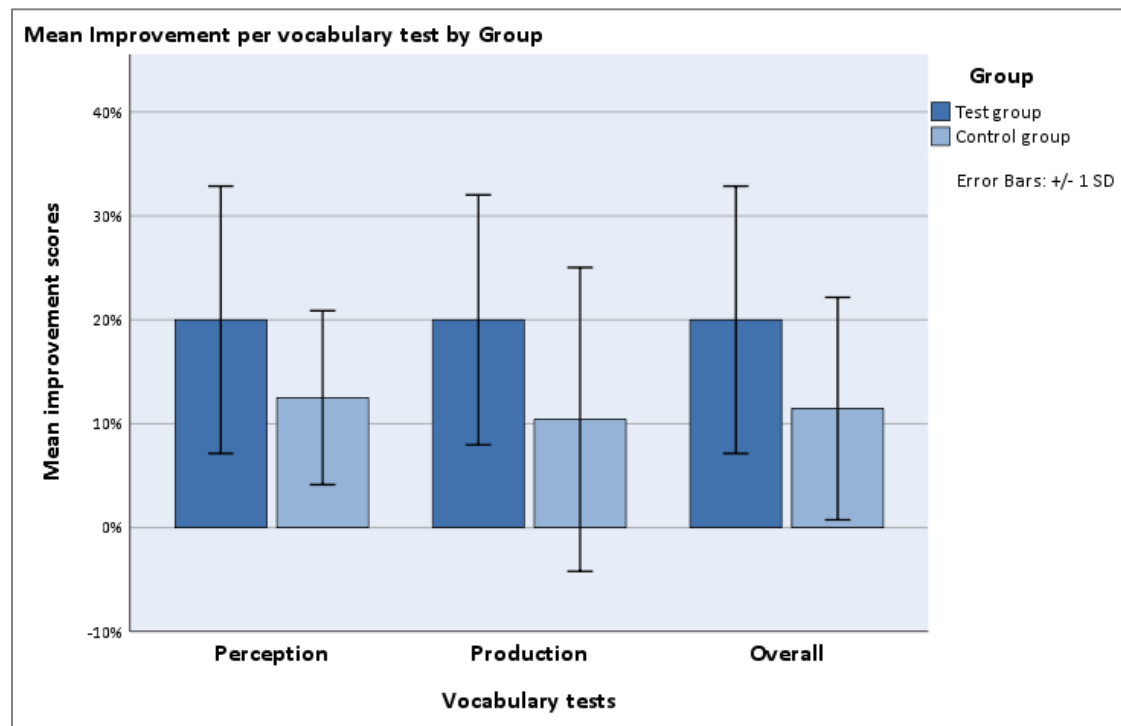


Figure 6: A Bar chart of the mean improvement scores per group, for the vocabulary perception test, the vocabulary production test and the overall improvement. Error Bars show 1 SD to indicate the large amount of individual variation.

### 3.2 Exploring assumptions

To ensure a parametric test (the MANOVA) was the correct option, it was necessary to check if the data for the independent variables (improvement scores) met the assumptions for parametric tests. First, to see if the data is normally distributed, we looked at skewness and kurtosis. This was checked through frequency z-scores and followed up with the Kolmogorov-Smirnov test. Additionally, the homogeneity of variance was checked, using Levene's test.

While looking at the overall frequencies of the data, including skewness and kurtosis of the variables "overall improvement", "production test improvement" and "perception test improvement", it turned out that the z-scores for the skewness and kurtosis for the variable "perception test improvement" showed both significant positive skew ( $p < .05$ ) and significant positive kurtosis ( $p < .05$ ).

However, a follow-up check with the Kolmogorov-Smirnov test showed that the improvement for the vocabulary perception test for the test group,  $D(5) = .30$ ,  $p = .161$ , and the control group,  $D(6) = .225$ ,  $p > .2$ , were both insignificantly non-normal, and thus we can assume that, despite the z-scores, the assumption of normality was met for all variables.

Levene's test for homogeneity of variance showed no significant differences in variance between the test group and the control group for the overall improvement,  $F(1,9) = .065$ ,  $p > .05$ , the improvement on the production test,  $F(1,9) = .084$ ,  $p > .05$ , and the improvement on the perception test,  $F(1,9) = .55$ ,  $p > .05$ . This means that the assumption for homogeneity of variances was met.

### 3.3 Running the MANOVA

Since the assumptions for parametric tests were met, a MANOVA was conducted to look at the differences between the test group and control group, for the improvement on the production test, the improvement on the perception test, and the overall improvement. Using Pillai's trace, no significant difference between the test group and control group was found,  $V = .14$ ,  $F(2,8) = .652$ ,  $p = .547$ ,  $\text{partial } \eta^2 = .14$  (large effect size).

This was unexpected, since the test group showed a larger mean overall improvement score ( $M = 20.0$ ,  $SE = 12.8$ ) compared to the control group ( $M = 11.5$ ,  $SE = 10.7$ ). However, a follow-up univariate ANOVA showed that this difference is not significant,  $F(1,9) = 1.45$ ,  $p = .259$ ,  $\text{partial } \eta^2 = .14$  (large effect size).

For the separate production and perception tests, two follow-up univariate ANOVA's were conducted. Although the test group also showed a larger mean improvement score for the vocabulary production test ( $M = 20.0$ ,  $SE = 12.0$ ) compared to the control group ( $M = 10.4$ ,  $SE = 14.6$ ), this difference was not significant,  $F(1,9) = 1.37$ ,  $p = .272$ ,  $\text{partial } \eta^2 = .13$  (moderate-large effect size). Again, for the vocabulary perception test, the test group shows a larger mean improvement score ( $M = 20.0$ ,  $SE = 15.1$ ) compared to that of the control group ( $M = 12.5$ ,  $SE = 8.4$ ), but this difference was also found to be insignificant,  $F(1,9) = 1.09$ ,  $p = .323$ ,  $\text{partial } \eta^2 = .11$  (moderate effect size).

## 4 Discussion

This section will discuss the results of the experiment and data analysis, including factors that were not analysed in the current study but might have been of influence. Next, it will discuss the play session observation findings and implications of these findings for the larger scale study. After that, it will mention how the experimental procedures and test materials can be improved before moving forward to the larger scale study. Finally, some suggestions for other future research will be discussed.

### 4.1 Experiment results explanation

The aim of the current study was to provide evidence for the effectiveness of the Noplica playhouse game Dancefloor. As such, we hoped to find significantly higher improvement scores for the test group that played the Dancefloor game, compared to the control group that played the Granny game. This would indicate that the Dancefloor game is an effective method of teaching Dutch vocabulary to immigrant children. Even though the conducted follow-up analyses revealed that there were no significant differences between the mean improvement scores for each group, they did show large effect sizes. This indicates that the direction of the attested differences is valid. The lack of significance could possibly be a result of the small group of participants, showing large individual variation in their vocabulary test scores. If the group sizes can be increased, it would become possible to see past this individual variation and more clearly compare the playhouse games effect. Hence, the large effect sizes provide optimism that the larger scale study will show a significant difference.

As has been mentioned before, the original plan was to have the control group play the Energy Center game, and not the Granny game. Because of this, the vocabulary tests include words that are present in both the Granny game and Dancefloor game. It is safe to assume that the experiment results will be more accurate when these words are replaced in the vocabulary tests, since it then solely focusses on the effects of the Dancefloor game. If there is no interference from the same words being learnt by the control group playing the Granny game, it is more likely that significant differences can be found. The clarity of the results might also increase if we can take some additional factors into account. Due to the current small group size, adding in possibly correlating factors gives a high chance of violating parametric test assumptions. But since these types of tests are robust to violations when the sample size is big enough, these issues are easily solved by using a sufficient number of participants. One of the factors that might correlate with the individual differences in improvement scores, is the length of stay in the Netherlands. This factor also showed large individual variation, which could have been enhanced by the COVID-19 pandemic, which coincided with this study (even delaying the research until the schools re-opened at the end of May 2020).

Since compulsory education applies to children from age five and up in the Netherlands, the participating children would have been enrolled in primary school almost immediately upon entry in this country. However, due to the lockdown and closing of schools caused by the COVID-19 Pandemic, it is unclear how much education these children actually received in the eight to twelve weeks prior to the start of the study. Especially for those children that had only been here for four months (so one month at most before the lockdown), setting up a home- or online-schooling alternative might have been difficult. This would indicate that there were some children participating without much prior knowledge of Dutch, while other children might have had at least a few months

of education here already. When the participants have more prior knowledge, one might expect the additional effects of the playhouse to be smaller, since their pre-test scores are expected to be higher. There are examples of this in the current study. For example, Table 1 shows that participant number 9 had only been in the Netherlands for four months and four days. The pre-test scores for this participant were the lowest (18.8% correct), but she also showed the largest improvement scores (41.3%). When we compare this to participant 7, who had been in the Netherlands for over a year and had already been going to school prior to the pandemic, we can see in Figure 4 (p. 21) that his pre-test score was very high, but he managed to decrease this score in the post-tests (most likely due to lack of focus). By controlling for the variation in length of stay in the analysis, we can see how much improvement score variation remains regardless of how long the children had been in the Netherlands.

Another useful factor could be the L1 of the participants. When this language is very similar to Dutch, it is relatively easy to score high on the vocabulary perception test. This was shown by participant number 3, who came from Germany, and was in the Netherlands for only four months and one day at the start of the study. All though she did not speak much Dutch, she was able to use her knowledge of German to score 61,9% correct on the vocabulary pre-tests. However, as she was able to understand so much Dutch already, her motivation to learn more might have been lower, since it was less necessary for successful communication. This might be why she obtained only around mean improvement score (18,1%). For the current study, L1 was considered when dividing the participants into groups, so that they would not be able to communicate in a shared L1 during the play sessions. For the larger scale study, this is still advisable. But possibly some factor of relatedness between the L1 and Dutch could be taken into account, to see if this might correlate with improvement scores. The results of that can then be used to determine if exploring this factor might be useful for future research projects, or if it is of little consequence when examining the effectiveness of learning through play.

## 4.2 Play session observations

In the previous chapter, the mean observation scores per group were reported, over all eight play sessions combined. However, just as with the vocabulary test scores, there was individual variation. Some participants were quite keen on chatting with the researcher during play sessions, resulting in high researcher involvement scores, but lower focus scores. Others were very enthusiastic and more focussed on the game, leading to lower researcher involvement. This seems to indicate that there could be some correlations between the different observations.

A different approach is to look at the observation checklist scores over time, instead of comparing groups. This would show if participants became more or less excited and focussed over time, if they repeated more or less words, and if they involved the researcher more or less for each session<sup>7</sup>. Finding correlations between the different observations over time might provide further insights into which aspects of learning through play might be more crucial, or if it depends on the type of game what aspect might be most influential. This was, in fact, the main point of scoring these observations: To see how these factors might influence the vocabulary test improvement scores. As Tomlinson and Masuhara (2009) indicated, being involved in a positive manner in an activity, facilitates learning. As

---

<sup>7</sup> For more details regarding this approach, including graphs of the different measures compared over time, please contact the researcher.

such, it is expected that higher levels of enthusiasm and focus would correlate with higher improvement scores. As has been explained previously, adding such factors to the analysis of the current study would cause higher chances of violating parametric test assumptions. Complicating the analysis might also lead to loss of power in such a small dataset. For the larger scale future study this would most likely not be the case, so it is recommended to see if and how the observation checklist scores might correlate to the improvement scores.

### **4.3 Suggestions for improvement of materials and procedure**

For the most part, the materials and procedures worked as expected. However, there is always room for improvement. First and foremost, the vocabulary tests need to be adapted, excluding those words that are also present in the Granny game without losing the balance in the item-distractor pairings and the overall balance between easy and more difficult items. A list of all overlapping items can be found in Appendix K: Overlapping words. It might also be a good idea to check with teachers at the beginning of a new study, to see which words will come up during classes in the test period, so these might also be excluded. This will yield even more concise improvement scores.

The method used for the vocabulary test seemed to work well, the children were very willing to participate and stayed motivated even when they did not know the correct response. For the production test the children sometimes used gestures and other tactics to point out that they did recognise the items that were pictured, indicating that they did understand the “task” the researcher presented; making sure that the Playhouse game images were “clear enough”. For the post-tests, most participants were quite excited when they recognised the test materials, to show how much they had learned since the pre-tests. However, there were a few participants that had previously scored very high and indicated that the perception test was perceived as boring. An additional challenge could be added for these children, by asking them to take the test as fast as they can. This will make it less boring for them, without having to adjust the difficulty levels or making it very different from the other participants; it will still show how many words they recognise.

Observations showed that over the weeks, the children became less interested in the game they were supposed to play, wanted to explore the other games more and more, and were also starting to question why it was always the same children that were taken from the class. These issues could be solved by extending the time period for the study so that there would be fewer sessions per week (giving the children more chances to participate in the normal class and thus maintain more interest in the playhouse), and having entire classes participate as opposed to just a small group per class. Conveniently, this automatically increases the overall number of participants, which is a vital point for the larger scale study.

Finally, the playhouse itself requires an upgrade. During the play sessions, the researcher noticed several small mistakes in the translations and/or the programming of the game. Additionally, part of the Dancefloor game did not function as it was supposed to. These problems were equal over all groups and the researcher responded to them similarly every time they occurred, so they will have had little effect on the current study. However, the researcher made a complete list of all issues, which will be sent to the Noplica team, and hopefully all errors can be corrected and all problems can be fixed before the start of the larger project. This list is included in Appendix L: Playhouse upgrades.

#### 4.4 Suggestions for other future research

One thing that became clear during the play sessions was that the different games elicit different types of play and different types of responses. This is to be expected, since they are very different games, but the variations in behaviour can lead to different expectancies. The Dancefloor game focuses on word learning, but does not provide much context and does not really require players to repeat any of the words. However, observations showed that when participants were more enthusiastic, they repeated more words as they were singing and dancing along to the songs. In this regard, it seems that higher levels of enthusiasm entice higher word repeating scores which would have a double impact on word learning. For the Granny game, the opposite effect seems to occur; when players became more enthusiastic in their game, they were less focussed on repeating the required sentences as they were asked, decreasing the word repeating score. Their focus was more on making funny noises or just screaming. One player even experimented with remaining quiet and seeing how the game would respond. Since every kind of response is counted as “correct” in the game, children would be more excited about these noise experiments than about actually playing the game and following the storyline. Hassinger-Das et al (2016) would expect that the Granny game would facilitate learning more than the Dancefloor game, since it provides a story as context to the words and thus leads to a deeper kind of understanding. But higher levels of excitement when playing this game would take the focus away from the story and context, indicating that more excitement for this particular game would actually take away from how much a player would learn. Taking these factors into account, a different line of study that looks more closely into these different approaches for learning through play can be explored by comparing the Dancefloor game to the Granny game.

One factor that the current play session observations did not record, was the level or amount of movement that the children had during the sessions. This could be another interesting thing to look at, since Kirk and Kirk (2016) found that increasing the amount and intensity of physical activity during lessons correlated with an increase in scores for their linguistic measurements. This indicates that moving more during the play sessions might facilitate learning even more. For the current study, it would be obvious that children playing the Dancefloor game would move around more than the children playing the Granny game, simply because of the different gameplay. To look more closely at how the amount of movements might influence improvement scores, a better method might be to look at just one game, and note how much the individual players move around during the play sessions. Children that are more active and moving can then be compared post hoc to those that moved around less. This method requires a large enough group of participants playing the same game, and might thus be more suitable for a different study that does not compare two games.

Another interesting point of comparison that has not been explored in the current study, is the role of the feedback present in the games. Since Weisberg et al (2013), Hassinger-Das et al (2016) and Hopkins et al (2019) agreed that guided play is the most beneficial form of play, the question is raised whether the playhouse provides this type of play. Is the feedback that a player receives from these interactive games close enough to what an adult guiding the activity might provide? In the Dancefloor game, the player receives positive feedback when pressing the correct buttons, but receives no feedback when a mistake is made. This might be quite similar to how an adult could respond, although they might also point out something is incorrect, especially when they are aware that the child does know the correct answer. The Granny game provides very different feedback.

When an incorrect choice is made by turning the wheel, the question is simply repeated until the correct answer is selected, which could be similar to an adult approach. However, when the player is asked to repeat a phrase, they always receive positive feedback, regardless of saying nothing or just making strange noises. This makes it less comparable to guided play, since an adult would not praise these kinds of responses. Apart from discussing whether the playhouse games are close enough to guided play, it might be interesting to see if additional adult guidance with the gameplay might yield better results. For additional projects, the researcher could be more closely involved with a subgroup of players, allowed to provide direct feedback and guidance while the children are playing. This subgroup could then be compared to one with minimal researcher involvement. Exploring the difference between this kind of feedback provided by a present adult to the type of feedback provided by the games themselves could point out whether the interactive gameplay is sufficiently close enough to guided play, or whether it might be improved so that the Noplica language playhouse can become the most optimal playing and learning environment for children.

## 5 Conclusions

The aim of this study was to provide evidence for the effectiveness of the Noplica language playhouse game Dancefloor. In order to obtain this, vocabulary tests were developed and an experiment was set up. Two groups completed the vocabulary tests twice, participating in eight fifteen minute play sessions between the two test moments. The test group played the Dancefloor game, and the control group played the Granny game. Their improvement scores on the vocabulary tests were calculated by subtracting the pre-test scores from the post-test scores, and then compared. Although the differences in improvement scores were insignificant, the test group showed a larger mean improvement, with a medium to large effect size. The developed experimental method and procedures proved quite successful. This study thus provided a good starting point for a larger scale study. When taking the suggested improvements into account, it is likely that significant results will be found. This will then hopefully show that children do significantly increase their Dutch vocabulary through playing the Dancefloor game in the Noplica playhouse, providing evidence that learning through play can be an effective method to facilitate language learning. Doing this will create opportunities for (immigrant) children to advance in school, and give them their best chance in life.



## 6 References

- Bustamante, A., Hassinger-Das, B., Hirsh-Pasek, K., & Golinkoff, R. (2019). Learning Landscapes: Where the Science of Learning Meets Architectural Design. *Child Development Perspectives*, 13(1), 34-40. DOI: 10.1111/cdep.12309
- Centraal Bureau voor de Statistiek. (2020, 20 april). *Hoeveel immigranten komen naar Nederland?* <https://www.cbs.nl/nl-nl/dossier/dossier-asiel-migratie-en-integratie/hoeveel-immigranten-komen-naar-nederland->
- Delle Luche, C., Durrant, S., Schmandt, S. & Floccia, C. (2015). A methodological investigation of the Intermodal Preferential Looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development*. 40. 151-172. DOI: 10.1016/j.infbeh.2015.05.005.
- Dunn, Lloyd M. & Dunn, Leota M. (2005). *Peabody Picture Vocabulary Test-III-NL, Nederlandse versie door Liesbeth Schlichting*. Harcourt Assessment B.V., Amsterdam
- Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, 27(1), 4-9.
- Hassinger-Das, B., Ridge, K., Parker, A., Golinkoff, R. M., Hirsh-Pasek, K., & Dickinson, D. K. (2016). Building vocabulary knowledge in preschoolers through shared book reading and gameplay. *Mind, Brain, and Education*, 10(2), 71–80. DOI: 10.1111/mbe.12103
- Hopkins, E., Toub, T., Hassinger-Das, B., Golinkoff, R. M. & Hirsh-Pasek, K. (2019). Playing for the future: Redefining early childhood education. In D. Whitebread et al. (Eds.), *The SAGE Handbook of Developmental Psychology and Early Childhood Education*. SAGE Publications Ltd.
- Kirk, S.M. & Kirk, E.P. (2016), Sixty Minutes of Physical Activity per Day Included Within Preschool Academic Lessons Improves Early Literacy. *J School Health*. 86. 155-163. DOI: 10.1111/josh.12363
- Singh, A., Uijtdewilligen, L., Twisk, J., Mechelen, W. & Chin A Paw, M. (2012). Physical Activity and Performance at School: A Systematic Review of the Literature Including a Methodological Quality Assessment. *Archives of pediatrics & adolescent medicine*. 166. 49-55. DOI: 10.1001/archpediatrics.2011.716.
- Tomlinson, B. & Masuhara, H. (2009). Playing to Learn: A Review of Physical Games in Second Language Acquisition. *Simulation & Gaming XX*. 40. DOI: 10.1177/1046878109339969.

## 7 Appendices

### 7.1 Appendix A: Dancefloor content

#### Sceneries and words

Scenery/ Theme	1. Zoo / Animals 1	2. Zoo / Animals 2	3. Farm / Animals 3	4. Smal Animals / Animals 4
1	vogel	beer	eend	bij
2	haai	neushoorn	varken	vis
3	vlinder	muis	schaap	mier
4	krokodil	schildpad	kip	vlieg
5	olifant	zebra	koe	wesp
6	tijger	wolf	paard	egel
7	ijsbeer	kameel	hert	slak
8	kikker	geit	gans	mol
9	aap	giraf	vos	mug
10	slang	leeuw	ezel	spin
Scenery/ Theme	5. Paint palette / colours	6. Body / Body parts	7. Classroom / Numbers	8. Street / vehicles
1	zwart	arm	één	vrachtauto
2	blauw	oog	twee	tractor
3	bruin	voet	drie	tram
4	groen	vinger	vier	vliegtuig
5	grijs	hand	vijf	fiets
6	oranje	hoofd	zes	boot
7	roze	mond	zeven	bus
8	rood	neus	acht	auto
9	geel	tong	negen	motor
10	wit	tanden	tien	trein
Scenery/ Theme	9. Kitchen / Food and drink 1	10. Vegetables and meat / Food and drink 2	11. Dining table / Food and drink 3	12. Table with fruit / Food and drink 4
1	appel	komkommer	rijst	mandarijn
2	banaan	sla	spaghetti	meloen
3	sinaasappel	champignon	melk	druiven
4	peer	paprika	water	ananas
5	brood	vlees	koekje	aardbei
6	worst	wortel	thee	citroen
7	boterham	bonen	boter	kers
8	ei	ui	chocolade	framboos
9	soep	spinazie	koffie	kiwi
10	kaas	tomaat	aardappel	pruim
Scenery/ Theme	13. Wardrobe / clothes	14. Family / People 1	15. Occupations / People 2	16. Table / Small household items 1
1	hoed	baby	kok	kom
2	jas	broer	kapper	lepel
3	sjaal	papa	politie	vork
4	shirt	oma	brandweer	bord
5	schoen	opa	tandarts	kop(je)
6	slipper	mama	bakker	mes
7	sok	zus	dokter	fles
8	trui	konijn	slager	pan
9	broek	hond	meester / juf	glas
10	riem	kat	boer	deksel

Scenery/ Theme	17. Home / Small household items 2	18. Play room / Toys	19. Map / Places to go	20. Room / Furniture and rooms
1	doos	bal	school	bed
2	telefoon	pen	winkel	stoel
3	zeep	blok	park	deur
4	tandpasta	pop	speeltuin	keuken
5	tandenborstel	puzzel	ziekenhuis	tafel
6	spiegel	potlood	zwembad	TV
7	klok	papier	strand	kast
8	kussen	knuffel	berg	wc
9	schaar	boek	boerderij	lamp
10	prullenbak	klei	restaurant	bank

## Sentences

	<b>Introductory sentences</b>
1	We zijn in de dierentuin. Hier wonen veel dieren.
2	Kijk! We zijn in de dierentuin. Ken jij alle dieren?
3	We zijn op de boerderij. Ken jij alle dieren?
4	Kom! Laten we naar buiten gaan. Kan jij de kleine dieren vinden?
5	Wat een mooie kleuren. Ken jij ze allemaal?
6	Dit is mijn lichaam. Weet jij hoe alle delen heten?
7	Kom, we gaan naar de klas! Ken jij alle getallen?
8	Laten we naar buiten gaan. Kijk, allemaal voertuigen!
9	Hier is de keuken. Heb je honger? Kijk naar al dat eten!
10	Van groente en vlees word ik sterk. Ik vind het allemaal lekker, jij ook?
11	Het is etenstijd. Ik heb zo'n honger. Wat zullen we eten?
12	Ik hou van fruit. Fruit is zo lekker. Hou jij van fruit?
13	Dit is mijn klerenkast. Mijn kleren liggen hier.
14	Ik heb een grote familie. Ik gal ze aan je voorstellen.
15	Zie al deze mensen! Wat doen ze voor werk? Hoe noem je hen?
16	Het is tijd om te eten. Wil je de tafel even dekken?
17	Ik ben net wakker. Ik moet me klaarmaken voor school. Wat heb ik daarvoor nodig?
18	Dit is mijn kamer. Mijn speelgoed ligt hier. Wil je met me spelen? Kijk, dit is mijn speelgoed.
19	Ik ga soms een dagje weg met mijn papa en mama. Waar ga jij graag heen?
20	Welkom in mijn huis. Er staan veel meubels. Ken jij ze allemaal?
	<b>Question sentences</b>
1	Waar is de/het [doelwoord]?
2	Kan jij de/het [doelwoord] vinden?
3	Kies de/het [doelwoord]!
4	Zie jij de/het [doelwoord]?
	<b>Feedback sentences</b>
1	Ja, je vond de/het [doelwoord]!
2	Het is je gelukt! Je vond de/het [doelwoord]!
3	Goed gedaan! de/het [doelwoord]!
4	Te gek! Je vond de/het [doelwoord]!
5	Goed zo! de/het [doelwoord]!
6	Prima! Je vond de/het [doelwoord]!

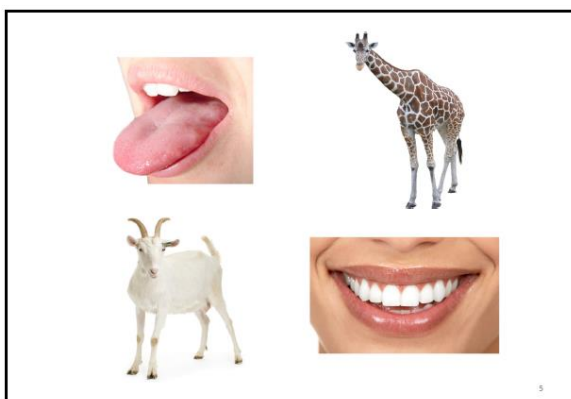
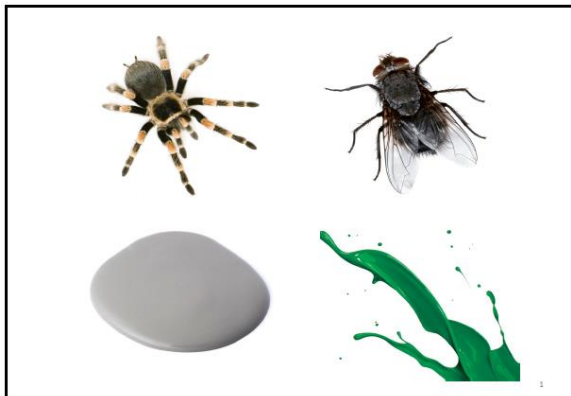
## 7.2 Appendix B: Granny stories

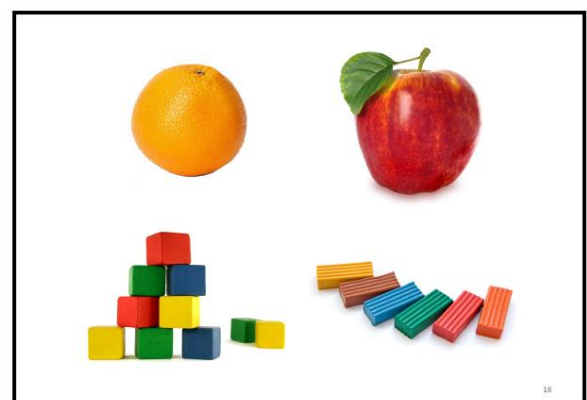
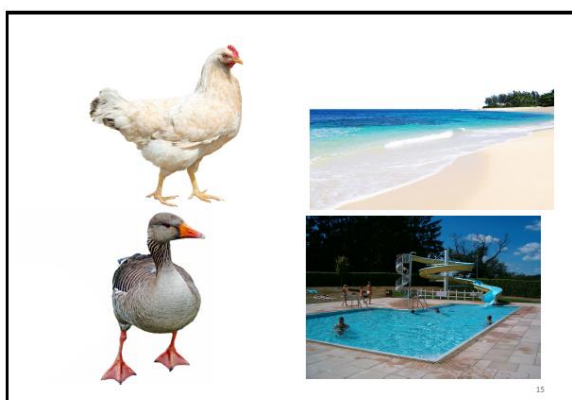
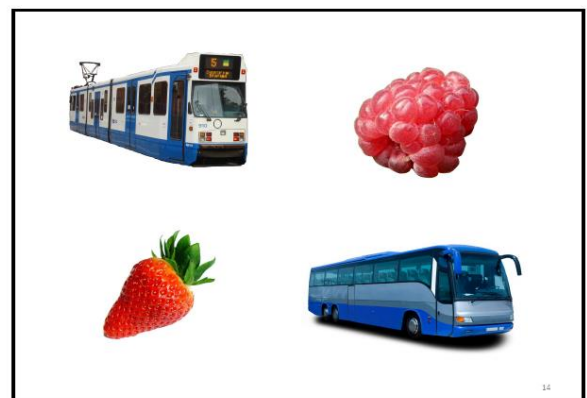
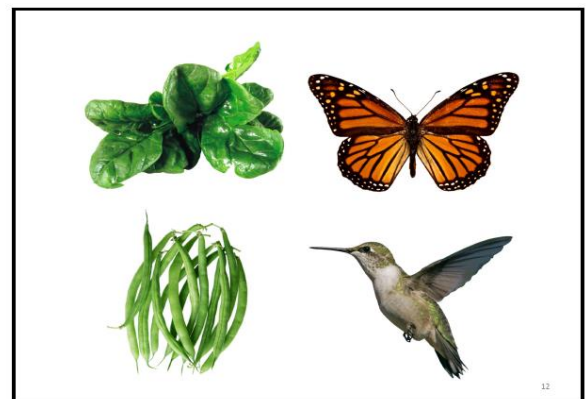
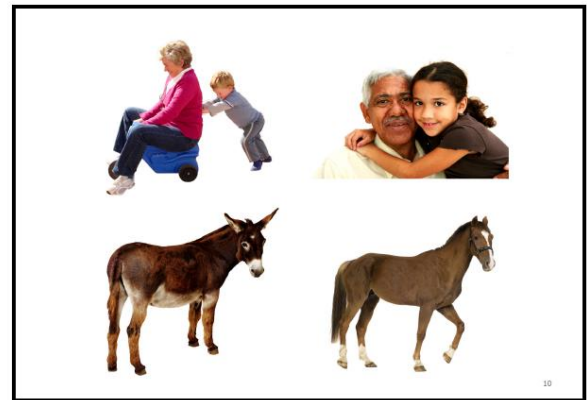
Dutch title	Translation	Content summary
1. Aankleden koud	Getting dressed – cold	Birdy and Grandma are going outside. The weather is cold. They need to pick the right items to wear. When they get outside, they build a snowman.
2. Aankleden warm	Getting dressed – warm	Birdy and Grandma are going outside. The weather is warm. They need to pick the right items to wear. When they get outside, they go buy ice cream.
3. Banen	Jobs	Grandma and Birdie go to the city. They encounter many people doing their job: a cab driver, waiter, fireman and a policeman.
4. Buren ontmoeten	Meeting neighbours	A new family moves in next door. They come over to introduce themselves: Rik, mother and father. Birdie and Rik play together.
5. Camping	Camping	Grandma and Birdie go camping together. They need to pick the things they need: a food, drinks, sleeping bags and a big tent. They use a hammer when putting up the tent and an axe to chop wood.
6. Gevoelens	Feelings	Rik is feeling sad. He is crying because he lost his teddy bear. Birdie and Grandma help find it. But the teddy bear hurt its ear. They put a bandage on it and go eat ice cream. Rik is happy again.
7. Jezelf wassen	Wash yourself	Birdie has played outside and got dirty. Grandma helps him wash. They point out different body parts to clean: belly, legs and feet.
8. Kaart	Postcard	The postman delivers a postcard. It is from Rik. There is a stamp on the card. Rik writes that he is on holiday. Yesterday he went swimming and today he is going to see a castle. Birdy writes a letter to Rik.
9. Kamer opknappen	Fixing up room	Birdie and Grandma have a new house. They mention different rooms and furniture. They want to make Birdies bedroom more cosy. They pick out a colour for the walls, new bedding and a poster for the wall.
10. Lunch maken	Making lunch	Grandma and Birdie are making lunch in the kitchen. Birdie wants to make a sandwich. They need bread, butter, tomatoes, lettuce, cheese, and a knife.
11. Naar de dierentuin	Going to the zoo	Grandma and Birdie go to the Zoo. They take the bus and buy a ticket. In the zoo, they see an elephant or a rhino (the player picks). Grandma loses her hat, and Birdie finds it.
12. Naar de dokter	Going to the doctor	Birdie is feeling sick. Grandma takes him to the doctor. The doctor draws some blood, looks at Birdies throat and takes his temperature. He prescribes some pills. The nurse brings them.
13. Onder water	Under water	Grandma and Birdie are swimming in the sea. They use diving goggles to look under water. They see fish and a turtle. A shark comes by and scares the fish. They hide behind rocks. One rock falls on the turtle. Birdie pushes the rock to free the turtle.

14. Op het strand	On the beach	Grandma and Birdie are at the beach. Birdie wants to build a sand castle. They also find seashells. They go swimming, and a squid ends up on Birdies head.
15. Schatzoeken	Treasure hunt	Grandma finds a treasure map. She takes Birdie to go find the treasure. They need to find the river and cross it. Then they have to go to a mountain. The treasure is in the forest, near a tree with seven monkeys. They find a chest with a crown in it.
16. Spelen in het park	Playing in the park	Grandma and Birdie go to the park. They go by bike. Birdie picks flowers for Grandma. He also jumps rope. It rains. Grandma brought an umbrella. They feed some bread to the ducks.
17. Spelen	Playing	Rik has come over to play. He wants to play football. They lose the ball behind a tree. Rik finds it again. Then they go play hide and seek in the house. Birdie hides inside a block tower. Rik hides behind the couch. When both are found, they go and drink lemonade.
18. Teken	Drawing	Birdie is bored. Grandma suggests drawing. They need crayons. Birdie wants to draw a castle. He needs to make a big square with 3 small squares on top. Then he wants to put a happy face on it. They draw a triangle as a nose. A yellow circle becomes the sun. Birdie gives the drawing to Grandma.
19. Theefeestje	Tea party	Birdie wants to have tea with his doll, his bear, and Grandma. Grandma pours the tea. It is hot. They also eat a cookie.
20. Verjaardagsfeest	Birthday party	It's Birdies birthday. He invited three friends. They all wear party hats. They have a cake with six candles on it. Birdie blows out the candles. The friends have brought presents.

### 7.3 Appendix C: Vocabulary perception test

*An overview of the passive vocabulary test flip-book pages*











## 7.4 Appendix D: Scoring sheets

Version 1:

# Scorekaart

Versie 1 | Participant:

Datum:

Test:

### Deel 1: Perceptie

Pagina	Item
1	Spin
2	Kop
3	Beer
4	Bergen
5	Tong
6	Tandenborstel
7	Aardappel
8	Egel
9	Mes
10	Oma
11	Koffie
12	Vlinder
13	Zeven
14	Bus
15	Zwembad
16	Appel
17	Kikker
18	Tandarts
19	Tractor
20	Hoed
1	Groen
2	Tomaat
3	Potlood
4	Bed
5	Geit
6	Oranje
7	Hand
8	Spiegel
9	Zes
10	Ezel
11	Deur
12	Bonen
13	Brandweer
14	Aardbei
15	Kip
16	Blok
17	Sok
18	Peer
19	Kat
20	Citroen

### Deel 2: Productie

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Version 2:

# Scorekaart

Versie 2 | Participant:

Datum:

Test:

## Deel 1: Perceptie

Pagina	Item
1	Groen
2	Tomaat
3	Potlood
4	Bed
5	Geit
6	Oranje
7	Hand
8	Spiegel
9	Zes
10	Ezel
11	Deur
12	Bonen
13	Brandweer
14	Aardbei
15	Kip
16	Blok
17	Sok
18	Peer
19	Kat
20	Citroen
1	Spin
2	Kop
3	Beer
4	Bergen
5	Tong
6	Tandenborstel
7	Aardappel
8	Egel
9	Mes
10	Oma
11	Koffie
12	Vlinder
13	Zeven
14	Bus
15	Zwembad
16	Appel
17	Kikker
18	Tandarts
19	Tractor
20	Hoed

## Deel 2: Productie

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

## 7.5 Appendix E: Vocabulary production test

*An overview of the vocabulary production test: Item distribution and test setup*

Table "easy" (green) and "difficult" (red) item distribution

Brood	Kok	Auto	Zebra
Olifant	vijf	Vinger	Lepel
Zwart	TV	Vis	Jas
Schaar	Knuffel	Baby	Chocolade



Figure: An image of the laminated A3 sized page used for the vocabulary production test. The set of cards used for this test contained the exact same images in the same size and were also laminated.

## 7.6 Appendix F: Test item selection brainstorm

### Brainstorm session

#### Suggestions for vocabulary tests

Passive/perception test: Two items per theme. Pages containing 2 targets and 2 distractors; each target-distractor pair from a different theme. Evenly distribute easy and difficult items.

Active/production test: 16 items. (20 themes; eliminate overlapping semantic categories: animals, food). Give the child cards; ask them to name which items they know and then put the cards on the correct spot. Again, evenly distribute easy and difficult items.

Perception test is easier and yields more data; provides more reliable results. Production test is more difficult so it needs to be more interactive and playful.

## Perception test items

Table: overview of selected items and distractors, and their classifications. Easy items are **green** (teacher questionnaire) or **yellow** (PPVT set 1-3)

Theme	Item 1	Distractor 1	Item 2	Distractor 2	Remaining easy
Dieren 1	Vlinder	Vogel	Kikker	Krokodil	Olifant aap
Dieren 2	Beer	Leeuw	Geit	Giraf	Schildpad
Dieren 3	Ezel	Paard	Kip	Gans	Koe eend varken
Dieren 4	Spin	Vlieg	Egel	Mol	Vis slak
Eten en drinken 1	Appel	Sinaasappel	Peer	Banaan	Brood worst
Eten en drinken 2	Bonen	Spinazie	Tomaat	Paprika	
Eten en drinken 3	Aardappel	Koek	Koffie	Thee	Water
Eten en drinken 4	Citroen	Meloen	Aardbei	Framboos	
Getallen	Zeven	Negen	Zes	Drie	Alles
Kleine huishoudelijke dingen 1	Kop	Kom	Mes	Fles	Glas
Kleine huishoudelijke dingen 2	Spiegel	Klok	Tandenborstel	Tandpasta	Doos schaar prullenbak telefoon
Kleren	Sok	Schoen	Hoed	Riem	Jas shirt broek
Kleuren	Groen	Grijs	Oranje	Rood	Blauw geel
Lichaamsdelen	Tong	Mond	Hand	Voet	Oog hoofd Neus
Mensen 1	Oma	Opa	Kat	Konijn	Baby papa mama hond
Mensen 2	Brandweer	Politie	Tandarts	Dokter	Juf*
Meubels en kamers	Deur	Kast	Bed	Bank	Stoel tafel lamp
Plaatsen om naartoe te gaan	Zwembad	Strand	Bergen	Park	School*
Speelgoed	Blok	Klei	Potlood	Pen	Bal boek
Voertuigen	Tractor	Motor	Bus	Tram	Auto fiets vliegtuig boot

\* Word/concept is classified as easy, but the image is unclear so it does not count as easy to name.

## Notes on distribution:

29 easy items | 6 item/distractor pairs | 9 distractors | 9 items | = 24 easy pairs | 16 difficult pairs  
 80 words total, 20 pages | 4 pages without difficult pair; those need 1 easy item per pair |

### Overview items/distractors Perception test

Page	Item 1	Distractor 1	Item 2	Distractor 2
1	Spin	Vlieg	Groen	Grijs
2	Tomaat	Paprika	Kop	Kom
3	Beer	Leeuw	Potlood	Pen
4	Bed	Bank	Bergen	Park
5	Tong	Mond	Geit	Giraf
6	Oranje	Rood	Tandenborstel	Tandpasta
7	Aardappel	Koek	Hand	Voet
8	Spiegel	Klok	Egel	Mol
9	Mes	Fles	Zes	Drie
10	Ezel	Paard	Oma	Opa
11	Koffie	Thee	Deur	Kast
12	Bonen	Spinazie	Vlinder	Vogel
13	Zeven	Negen	Brandweer	Politie
14	Aardbei	Framboos	Bus	Tram
15	Zwembad	Strand	Kip	Gans
16	Blok	Klei	Appel	Sinaasappel
17	Kikker	Krokodil	Sok	Schoen
18	Peer	Banaan	Tandarts	Dokter
19	Tractor	Motor	Kat	Konijn
20	Citroen	Meloen	Hoed	Riem

### Production test items

#### Remaining easy items (to select from: 42 items /8 items required)

Olifant	Aap	Schildpad
Koe	Eend	Varken
Vis	Slak	Brood
Worst	Water	Auto
Glas	Doos	Schaar
Prullenbak	Telefoon	Jas
Shirt	Broek	Blauw
Geel	Oog	Neus
Boot	Hoofd	Baby
Papa	Mama	Hond
Stoel	Tafel	Lamp
Bal	Boek	1
Fiets	Vliegtuig	2
4	5	8

### Selection

8 items (16 total | **8 easy** | 8 difficult) *only the easy items are present here*

Olifant	Vis
Brood	Vijf
Auto	Schaar
Jas	Baby

## Subsets easier items: From teacher questionnaire and PPVT

### *A: Teacher – know and say*

- Vogel
- Alle getallen (all numbers)
- Bus
- Auto
- Jas
- Schoen
- Baby
- Papa
- Mama
- Juf
- Bal

### *B: Teacher – know*

- |              |            |          |
|--------------|------------|----------|
| - Olifant    | - Aap      | - Geel   |
| - Groen      | - Rood     | - Hand   |
| - Oog        | - Voet     | - Neus   |
| - Hoofd      | - Mond     | - Banaan |
| - Fiets      | - Appel    | - Shirt  |
| - Brood      | - Water    | - Hond   |
| - Sok        | - Broek    | - Schaar |
| - Kat        | - Telefoon | - Boek   |
| - Prullenbak | - Potlood  | - Stoel  |
| - School     | - Zwembad  |          |
| - Deur       | - Blauw    |          |

### *C: PPVT lower sets (practice sets, set 1-3)*

#### Items:

- |             |             |
|-------------|-------------|
| - Schildpad | - Koe       |
| - Vlieg     | - Oog       |
| - Hand      | - Vliegtuig |
| - Broek     | - Baby      |
| - Kat       | - Schaar    |
| - Bank      |             |

#### Distractors:

- |           |           |           |
|-----------|-----------|-----------|
| - Vogel   | - Vlinder | - Olifant |
| - Giraf   | - Eend    | - Varken  |
| - Paard   | - Gans    | - Vis     |
| - Slak    | - Mond    | - Neus    |
| - Tractor | - Fiets   | - Boot    |
| - Auto    | - Appel   | - Banaan  |
| - Worst   | - Shirt   | - Riem    |
| - Hond    | - Kop     | - Glas    |
| - Doos    | - Klok    | - Boek    |
| - Bed     | - Tafel   | - Lamp    |

## 7.7 Appendix G: Consent Forms

### I) School consent form, original Dutch version<sup>8</sup>

Radboud Universiteit



#### INFORMATIEDOCUMENT EN TOESTEMMINGSVERKLARING SCHOOL

##### Uitleg onderzoek

Op het speelplein van uw school is een taalspeelhuisje geplaatst. Het doel van dit onderzoek is om te kijken in hoeverre het spelen in dit taalspeelhuisje bijdraagt aan het leren van Nederlandse woorden door de kinderen. Er wordt hierbij specifiek naar (de woorden uit) één spel gekeken: Dancefloor. We willen het onderzoek gaan uitvoeren bij 12 kinderen tussen de 4 en 7 jaar (uit de kleuterklassen en het wasberenwoud), verdeeld in kleine groepjes.

Door de onderzoeker is een korte taaltest ontworpen die specifiek vraagt naar de woorden in het spel 'Dancefloor'. Dit testje zal op twee momenten worden afgenomen, een keer voorafgaand aan de "speelfase" en een keer achteraf. De kinderen worden hiervoor individueel uit te klas gehaald. Het afnemen van de test duurt ongeveer 5 tot 10 minuten per kind. We nemen onze eigen materialen mee, en zouden graag gebruik maken van een aparte ruimte voor het afnemen van de test. Ook zouden we graag geluidsopnames maken van de antwoorden van de kinderen, om de scores te kunnen controleren en de uitspraakvarianten te kunnen onderzoeken. Uiteraard zullen we ons houden aan alle geldende maatregelen m.b.t. hygiëne en afstand i.v.m. het coronavirus. Het testmateriaal zal tussendoor steeds schoongemaakt worden.

Tijdens de speelfase zal de onderzoeker de leerlingen in groepjes van 3 meenemen naar het speelhuisje, om hier steeds een kwartier een van de spelletjes te gaan spelen. Deze sessies zullen zo ingepland worden dat alle kinderen gedurende 4 weken 2 keer per week kunnen spelen. Dit alles zal in overleg met de school en de leerkracht gebeuren. De leerlingen zullen verdeeld worden in 2 groepen (die later weer opgedeeld worden in groepjes van 3), die elk een ander spel in het taalspeelhuisje gaan spelen. De testgroepjes spelen Dancefloor, en de controlegroepjes spelen Energy Centre.

##### Informatievoorziening

De leerlingen zullen in de klas door de leerkracht en de onderzoeker geïnformeerd worden over het onderzoek. Indien nodig zullen de ouders/verzorgers ter informatie een brief ontvangen. Hierin staat dat de kinderen mee zullen doen aan een onderzoek, dat er opnames gemaakt zullen worden, en op welke wijze de ouders bezwaar kunnen maken als ze dit niet (meer) willen. De deelnemers zullen geen vergoeding krijgen voor het deelnemen aan het onderzoek.

##### Wat gebeurt er met de gegevens die we in dit onderzoek verzamelen?

De onderzoeksgegevens die we in dit onderzoek verzamelen, zullen gebruikt worden voor een Masterscriptie, en kunnen door wetenschappers gebruikt worden voor datasets, artikelen en presentaties. De anoniem gemaakte onderzoeksgegevens zijn ten minste 10 jaar beschikbaar voor andere wetenschappers. Als we gegevens met andere onderzoekers delen, kunnen deze niet tot de deelnemers herleid worden.

##### Nadere inlichtingen

Voor vragen of verdere informatie over het onderzoek kunt u contact opnemen met:

Uitvoerend onderzoeker; Elea Thijssen, email: [elea.thijssen@outlook.com](mailto:elea.thijssen@outlook.com), tcl: 0636073327

Verantwoordelijk onderzoeker; Paula Fikkert, email: [p.fikkert@let.ru.nl](mailto:p.fikkert@let.ru.nl)

<sup>8</sup> For a translation, please contact the researcher, E. Thijssen (contact details mentioned in the form)

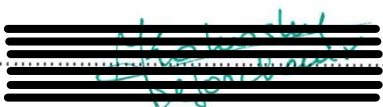


### Toestemming

Ik geef de onderzoeker(s) van de Radboud Universiteit toestemming het onderzoek uit te voeren op mijn school

Naam school: ..... de Bloemberg .....

Naam en functie: ..... Iris Kohasky Deforchaux, directeur .....

Handtekening: .....  .....

Datum: ..... 27-05-2020 .....



## ONDERZOEK OP SCHOOL / RESEARCH STUDY AT SCHOOL

**Naam onderzoek / title of research study:** *spelend taal leren / learning language through play*

**Verantwoordelijke onderzoeker / researcher responsible:** *Paula Fikkert*

**Uitvoerend onderzoeker / executive researcher:** *Elea Thijssen*

### ***Uitleg onderzoek (English follows Dutch)***

Een onderzoeker komt kijken of uw kind Nederlandse woorden leert van het taalspeelhuisje dat op het schoolplein staat. De onderzoeker gaat twee keer een testje/spelletje maken met uw kind. Hierbij wil de onderzoeker de stem van uw kind opnemen. Uw kind gaat ook met klasgenootjes in het taalspeelhuisje spelen.

De gegevens die de onderzoeker verzameld, zullen zonder de naam van uw kind (anoniem) worden gebruikt en bewaard.

Als u wilt dat uw kind stopt mee te doen met het onderzoek, stuur dan een e-mail naar de onderzoeker: [elea.thijssen@outlook.com](mailto:elea.thijssen@outlook.com), of zeg het tegen juf van uw kind. Alle verzamelde gegevens van uw kind zullen dan verwijderd worden.

### ***Verklaring ouder/voogd***

Als u niets laat weten, geeft u daardoor aan het goed te vinden dat uw kind mee doet aan het onderzoek. Uw kind doet vrijwillig mee, en hij/zij mag stoppen als hij/zij dat wilt.

### ***Explanation of Study***

A researcher is coming to find out if children can learn Dutch words by playing in the Language playhouse that is located at the school. The researcher will take a short test/game with your child, two times. During this test, the researcher wants to record your child's voice. Your child will also get to play in the playhouse, with his/her classmates.

Any data the researcher collects, will be stored and used without your child's name (anonymously).

If you do not want your child to participate in this study, please send an e-mail to the researcher: [elea.thijssen@outlook.com](mailto:elea.thijssen@outlook.com), or tell your child's teacher. All of the collected data from your child will then be removed.

### ***Statement of Parent/guardian***

If you do nothing, this means that you agree to let your child take part in the research study. Your child is taking part willingly, and he/she can stop at any time he/she wants to.

## 7.8 Appendix H: Observation Checklist

# Checklist spelobservatie

---

Datum:

Sessie:

Aantal kinderen:

(namen / groep / nummer)

Starttijd:

Eindtijd:

Duur sessie:

Evt. reden i.v.m. eerder afgebroken:

Interesse / betrokkenheid 1-2-3-4-5-6-7

Enthousiasme 1-2-3-4-5-6-7

Speelden ze zoals “bedoeld” met het spel: wel | niet

Opmerking (als niet, hoe dan):

Hardop herhalen woorden: niet | 1-2-3-4-5-6-7 | continue

Opvallend tijdens spel:

Opvallend individuele kinderen:

1]

2]

3]

4]

Mate waarin onderzoeker betrokken/aangesproken wordt: 1-2-3-4-5-6-7

Overige opmerkingen:

## 7.9 Appendix I: Colouring pages

*Images of the colouring pages the children received as a thank-you.*

*These images were created by the researcher and can be freely used again for any further study.*



An image of Birdy, the main character in the Granny game and Noplica logo

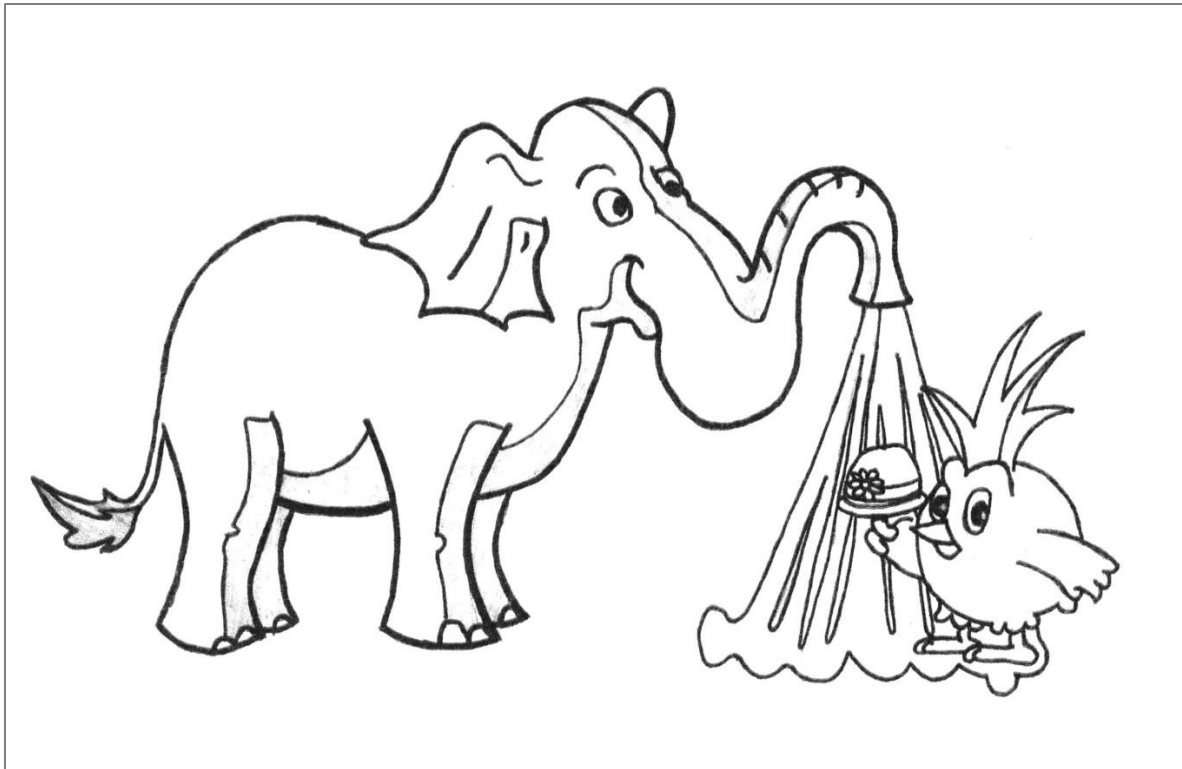


Figure: An image from the Granny story "Zoo"



Figure: An image from the Granny game story "Tea party"

## 7.10 Appendix J: Classroom questionnaire

*Teacher classroom questionnaire (In Dutch<sup>9</sup>). The researcher had filled out the names and added sufficient tables so that there was one for each child.*

### Vragenlijst: In de klas

Om een beeld te krijgen van de eventuele invloed die het uitvoeren van het experiment op het gedrag van de kinderen zou kunnen hebben, wil ik u vragen om deze vragenlijst in te vullen.

De eerste twee vragen hebben betrekking op het gedrag van de leerling, op de dagen dat er met de onderzoeker gespeeld is in het taalspeelhuisje. Omcirkel het antwoord. Het gaat hierbij om het effect dat u gemiddeld merkt. U hoeft dus niet voor elke sessie opnieuw de vragen te beantwoorden.

Mochten er dingen alleen op specifieke dagen opvallen, dan kunt u dit bij *opmerkingen* invullen.

Vermeld er dan de datum bij.

<b>Naam kind:</b>						
Het gedrag van het kind is:	<i>veel drukker</i>	<i>drukker</i>	<i>(geen verschil)</i>	<i>rustiger</i>	<i>veel rustiger</i>	dan op een normale dag
De concentratie/aandacht van het kind tijdens de lessen is:	<i>veel slechter</i>	<i>slechter</i>	<i>(geen verschil)</i>	<i>beter</i>	<i>veel beter</i>	dan op een normale dag
<b>Opmerkingen</b> (valt er iets op, qua gedrag of bijvoorbeeld taalgebruik)						
<b>Heeft u het idee dat het kind het onderzoek leuk vindt?</b> Waaraan merkt u dat?						

---

<sup>9</sup> Translations in English can be provided by the researcher upon request.

## 7.11 Appendix K: Overlapping words

*List of the words that occur in both the Dancefloor game and the Granny game.*

aap	acht	auto
bal	bank	bed
berg	blauw	blok
boek	boter	boterham
brandweer	brood	bruin
bus	dokter	drie
een	eend	fiets
geel	grijs	groen
haai	hoed	jas
kaas	kast	koekje
kop(je)	melk	mes
mond	negen	neushoorn
olifant	oma	oranje
park	pen	politie
pop	rood	roze
schaar	schildpad	shirt
sjaal	sla	stoel
strand	tafel	thee
tomaat	tong	twee
vier	vijf	vis
voet	wit	zeep
zes	zeven	zwart

## 7.12 Appendix L: Playhouse upgrades

*Noplica language playhouse: a list of defects, mistakes, and suggestions for a possible update of the Dancefloor game and Granny game.*

<b>Dancefloor</b>			
<b>Date</b>	<b>Theme</b>	<b>Defect/mistake</b>	<b>Suggestion</b>
June - July 2020	n.a.	Blue lights not lighting Right-hand panel, below the screen. Touch function still works.	
June – July 2020	Speelgoed	Song after first round still has audible English song. Plays simultaneously with Dutch words.	
June – July 2020	Kleuren	During the song/movie after the first round, the order of the colours shown doesn't match with the order in the song.	
June – July 2020	Cijfers	After the second round, in the clip/song: één, twee, "three". (English word instead of Dutch).	
June – July 2020	Beroepen	Picture: Juf. Word: Meester.	Both are supposed to be in the game. Make sure the picture matches the word.
June – July 2020	Eten en drinken 3 (eettafel)	After the first round, the English song is played with the Dutch words over it.	
June – July 2020	Lichaam	Picture: 1 tooth (molar). Word: "tanden"	Change the picture so that it shows more teeth.
June – July 2020	Lichaam	After the second round, not all the body parts are mentioned; the song/movie seems to skip some parts.	
June 19 2020	Eten en drinken 4 (fruit)	After part 1: "kiwi pruim kiwi pruim". So a short part is repeated instead of naming the correct fruits.	
June – July 2020	Dieren 4 (kleine dieren)	The picture/drawing of the hedgehog (used in the movie and second round) does not look like a hedgehog at all.	Find a better drawing of a hedgehog.
June – July 2020	Kleine huishoudelijke dingen 1 (tafel)	After the first round, the screens show the picture of "bord" (plate), but the word is "deksel" (lid).	
June 26 2020	n.a.	The weather was really hot: over 30° C. The left-hand panel did not always respond. The lights are off / do not flash. And the top-left button did not respond at all.	Most likely cause is high temperature. Not sure if there is a solution for this, maybe just not play the game in such hot weather to prevent damage?
July 1 2020	Eten en drinken 4 (fruit)	Sentence: "waar is de druiven"	Sentences expect singular. Maybe change to "druif"?



June – July 2020	Kleine huishoudelijke dingen 2	After the second round, the English words are still in the song that plays.	
June – July 2020	Beroepen	Image (photo): Farmer This is not a stereotypical Dutch farmer.	Find a new picture
June – July 2020	Plaatsen om naar toe te gaan	Images: boerderij, ziekenhuis, park, speeltuin, restaurant, school, winkel  These do not match typical Dutch examples of these places and are difficult to recognise (even by native Dutch speakers)	Replace these pictures

Granny			
Date	Story	Defect/mistake	Suggestion
June – July 2020	Dierentuin	Audio plays: “we moeten eerst een kaartje kopen”. Text on screen: “we moeten eerst een keertje kopen”	Change tekst to: “kaartje”
? (issue was explained by Erwin, is apparently recurrent)	n.a.	When the games are turned on in the morning, Granny game sometimes does not start up. Turning it off and on again fixes this.	
June – July 2020	Nieuwe burens	text: “dank jewel”.	Change tekst to: “dankjewel”.
June – July 2020	Kamer opknappen	Text: “kun je het deken met rondjes vinden”	Change to: “de deken” (het deken is more Belgian Dutch)
June – July 2020	Kamer opknappen	There is a poster of “meneer de uil”. Young children won’t get this reference	Consider changing it to a more modern reference
June – July 2020	Spelen	They are about to drink lemonade. Oma asks: “zie jij de ranja?” The pictures are of an orange teapot and a jug of lemonade. Some of the children mishear “ranja” as “oranje” and select the teapot. It overcomplicates.	Change the word “ranja” to “lemonade”  Or change the orange teapot to a different colour.