Transfer of Educational Skills from Games to Classroom tasks - A Case Study using iRead to Improve Reading and Writing

Kay Berkling
Cooperative State University
Karlsruhe, Germany
kay.berkling@dhbw-karlsruhe.de

Birgit Franken
Cooperative State University
Karlsruhe, Germany
Birgit.Franken@dhbw-karlsruhe.de

ABSTRACT

This paper describes a short-term study of employing learning games in a German school setting. Games focusing on long and short vowel identification are chosen from the suite of iRead Project games and presented to groups of first and second graders. The children are presented with a game of identifying long vowels seeing around 25 words in 1-2 minutes once a week for four weeks. Their writing skills are recorded before and after the intervention. Recordings made during game play are used to analyze their performance during game play with respect to correctness and speed. Reading skills practiced in the game are not always transferred to the children's writing skills. In general, the impact was greater for first grade children. Our findings will support future study designs for how games can be most supportive in classrooms.

ACM Classification Keywords

D.2.2. User interfaces: K.8.0. Games

Author Keywords

Educational Game; Education; Reading; Writing; Eye-Tracker; Performance;

INTRODUCTION

The origins for this paper lie in the question that poses itself naturally when taking the first steps of digitalizing the classroom. Digitalization has many meanings. In our context, we wanted to look at how learning games can be used in the classroom. In this publication we would like to share the learning path that we took but more importantly share our findings that will help in understanding where the gain of games in schools can be and how they can be employed in the classroom. The data and their evaluation or analysis can be considered a preliminary study with strong indications of which research directions should be pursued to see if the findings can be supported. As a result, we can strengthen the effectiveness of the games in the classroom and gain a fuller understanding of why and where they should best be employed. At the moment the German government is pushing into the direction that schools should start using computers and mobile devices in schools. For German language games applicable for school use in elementary school, not too much is known on how to choose educational games and how teachers effectively "orchestrate" their use within the classroom, a persisting issue in the field of serious games [21]. Furthermore, with respect to German language games in elementary school in particular, there are few known studies that look at the academic impact that serious language games have on the students [18, 3]. In order to prepare further work on recommendations for integrating games into the German elementary classroom, this paper focuses first on the academic impact of a particular game that trains a specific but central skill to reading and writing success [4]. Namely, practicing (with immediate feedback on correctness) the identification of short and long vowels and recording the impact on orthographic and reading skills.

We would like to formulate the following hypothesis: A game seeks to improve a particular skill in a player. We would like to show that repeated play will have an academic impact on that skill. Even though four trials over four weeks resulting in only a few minutes of game-play are very little, we expect to see some effect on reading and writing skills for a subset of the children as a result of the intervention. Understanding these effects more closely, will support future work on assigning games to grades and children as well as recommending teachers how to integrate the game within their lesson plan (for example: Whether the game should be played only after introducing the related concept, how many times a game is expected to be played by a particular child until skill mastery and whether and how the impact can be measured).

In order to share the process we have gone through for understanding the impact of the game on classroom skills, the paper is structured as follows. After a brief introduction into the EU-Project iRead, under which umbrella the described study was conducted, the design of the current pilot study and measurement of knowledge transfer to reading and writing skills are described. The results of the game play itself are reported, followed by findings on reading and writing skills. In conclusion, we will discuss to what extend these findings can be related to the game play. Finally, there is an outlook on the kind of work that remains to be done before these findings can

iRead is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 731724. The results presented reflect only the authors' view and the Agency is not responsible for any use that may be made of the information it contains.

In: J. Arnedo-Moreno, C.S. González, A. Mora (eds.): Proceedings of the 3rd International Symposium on Gamification and Games for Learning (GamiLearn'19), Barcelona, Spain, 22-10-2019, published at http://ceur-ws.org

be substantiated with enough data and applied to classroom integration.

IREAD PROJECT DESCRIPTION

The iRead games are developed for L1 and L2 learners of English. In addition the games are ported to German, Spanish and Greek.

Game Content

There are a variety of different game mechanics provided by the iRead game collection, such as sorting (combining letters to words) or decomposing (cutting of prefixes), choosing the correct answer among 3-5 options (see for example Figure 1) or true/false (see for example Figure 2). Some of the games have time-limits. Each game mechanic has several graphical realisations. For example, multiple choice is visualized as choosing a path across water, picking the correct rail-road track or moving the correct article to construct a path. The content is automatically supplied to the game using a rule-based word selection mechanism. The difficulty level is informed through a domain model that sequences a range of language prerequisites according to the school curricula. Words are presented in a sequenced manner of difficulty at the learner's level of phonology (the study of grapheme-phoneme correspondence), morphology (the study of prefix and suffix recognition and usage), and syntax (the study of endings within sentences, such as conjugation, and grammatical knowledge that requires syntax analysis).

Users

The game is built to support reading acquisition in L1 and L2 setting for young school children within the classroom setting. For the L1 setting, the applicability of the games to dyslexic children is also studied. The games are developed for Greek, Spanish, English and German. The iRead applications are deployed in the classroom setting to support teachers with their work. The teacher makes the final decision on how to incorporate the games and eReader. For example, the teacher can give learners the option of playing this software program or read on the eReader as two of several "learning stations" during a free study period.

Points and Feedback

Within a mini-game, points and penalties are given for correct and incorrect input. Feedback is therefore immediate when performing each particular task. If the player makes mistakes, specific hints can be given to help rethink the problem in a second attempt. The hints are task specific. For example, the learner sees the words 'Schule', 'sehen', 'spaten', 'stehen', and 'sagen'and the learner has to click on the words that contain the sound /sch/. When the user selects an incorrect word, the system gives the hint "the letter <s> can make the sound /sch/ when it comes before <t>". Then the user is given an additional time to select the correct words.

Subset of Games chosen

For the purpose of this study, two representative game mechanics were chosen because they provided the maximum content within the shortest amount of time. Their mechanics are (a) pick the correct answer out of an average of 5 choices without time-limit and (b) true/false with time limitation. Each game provides about 25 words in one playing session of about 1-2 minutes. The automatically generated content for the games was chosen according to the explanation in the next section.

		Participants		
		Eye-Tracker	Writing	Games
Grade (24)	1	12	20	17
Grade (23)	2	10	23	23

Table 1. Statistics of Participants in Study

DESCRIPTION OF STUDY

The set up of the study included four visits to a classroom. 10 classrooms participated in the study but for the purpose of this paper, only two were transcribed and evaluated. The numbers are listed in Table 1.

A Tiny Introduction to German Orthography

A "Trochee" is a metrical foot consisting of a stressed syllable followed by an unstressed one. This word pattern forms the basis of the core German word and its orthography [6]. Within this 2-syllable word, several patterns can occur, some of which create more difficulty for children than others. Some forms are easier to transform from graphemes to phonemes and vice-versa than others. For example, a relatively easy word to read and write would be "Hase" (rabbit) since the phoneme to grapheme transcriptions is fairly straight forward if the child knows the basic letters. "hasse" (hate) is more difficult because it requires understanding of the <ss> grapheme that shortens the preceding vowel. Similar to the English "cut" and "cute" lessons from Phonics, the "silent-e" is studied explicitly by learners. In German however, this pattern is not studied explicitly for beginning readers. The pattern holds true for all vowels: <a,e,o,u>, which are written the same way for long and short vowels. Their length depends on the following consonants. The long and short vowels <i,ie> are the exception, as their spelling directly reflects their length in regular words. A child that has not learned the orthographic principles will pronounce both "hasse" and "Hase" with a long vowel and will semantically map the word "Rosse" (horses) to "Rose" (rose). As a consequence, reading is painful for them and their spelling, which simply omits double consonants, hard to read. The games are configured to present a set of 2-syllable short vowels words and only one long-vowel word for them to find. The selection is automated and based on a 15.000 word dictionary. Since the dictionary still contains some mistakes, the kids are told that the computer can be wrong and sometimes there is no correct answer to choose from.[1]

Pre-Visit

Before the initial visit, the classroom is asked to take a written "exam" containing 10 words (1st Grade) or 20 words (2nd

[1] For example, "Koffer" is marked with a long vowel in the dictionary.

Grade). The following criteria are applied for choosing the words:

- The words are chosen from the childlex [19], taking half from high frequency words and half from the low frequency word list in order to avoid measuring memorization of words over skills of applying orthographic patterns.[2]
- All words have the form of “Trochee".
- The words are represented by a picture and a sentence with the missing word to support correct word choice.
- The spellings are analyzed for the two features that were trained with the game (see above).

Meta-data for each child is included in the data elicitation. However, this information is not relevant for the purpose of this paper.

Visit 1

The first visit starts with a small 1-1 introduction of an orthographic concept for each participant for the game play. The graphemes <ie> vs. <i> and their function within a word is explained. Examples of minimal pairs are given, which the children are asked to read: “Riese" (giant) vs. “Risse" (crack). In the beginning, with few exceptions, all children will read both of those words as giant. They mostly agree though that these have two different meanings and that it would be useful to be able to distinguish these while writing. “Wiese" (meadow) vs. “Wissen" (knowledge) is used as a second example. The second of the double consonants is named the magic letter (as it also cannot be heard - in analogy to the silent-e in English). No mention of the other vowels is made during this introduction. We do ask the kids to sound out what a short /ɪ/ would sound like. This is a new task to them, as they are only taught the long sound of the vowel. This introduction takes only five minutes, after which the children are asked to play their first game where they are asked to identify the long “ie” in one of the presented words. The game takes an average of only two minutes. The interface is shown in Figure 1 and is the same for Visit 2 and 3. Distractors are short vowels with double consonants or <tz>. As the kids play, they receive feedback from the game: Correct words are turned green on the chosen stone. Incorrect words turn red and the avatar jumps back to the previous stone. After a second wrong answer, the avatar automatically moves to the correct stone. An oral feedback is given by the system, saying that the short vowels can be recognized because of the double consonants.

Visit 2 and 3

The next two visits, the children are playing the game with different content. Whereas previously, they had to only find the long “ie", these games now include all the long vowels. Because the oral feedback after a mistake is barely audible with the music, the system feedback, that a double consonant denotes short vowels, is repeated by the researcher sitting with the child. We also remind them, that spotting an <ie> is the easiest way to recognize a long vowel. The child is reminded that the computer could be making a mistake also. Often kids will be excited to recognize these mistakes. A screen recorder records all sessions again.

[2]Analyzing the performance differences between these two groups is another study and out of scope for this publication

Figure 1. Screen-shots of slow games, finding long vowel <ie>, or with alternative content, finding any long vowel (“ie, e, ö”-sequence is depicted here)

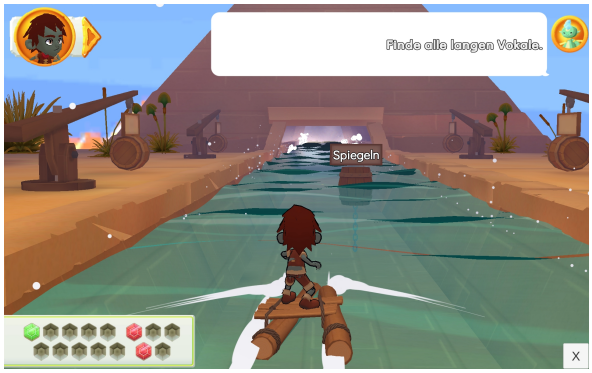

Figure 2. Screen-shot of fast game, where player has to hit the correct words in time

Visit 4

During the final visit the children will repeat the game from sessions 2 and 3. During this game they are asked to tell the researcher how they know whether they did something correct or wrong during the play. Next, they are allowed to play a fast game if they wish, which they all did. In this speed game, limited time is available for the player to choose words with long vowels. A screenshot of this game is provided in Figure 2.[3]

Post-Visit

A post-test is distributed to the classroom. Different words following the same patterns of word selection are presented to the children.The spelling errors with respect to the features practiced in the game are then analyzed: Correct spelling for short and long vowels using the “magic” consonant and the correct usage of <ie>.

[3]Analyzing results from the fast game is beyond the scope of this paper.

	1	2	3	4	5
S	100%	100%	63%	86%	90%
R1	88%	100%	86%	88%	100%
R2	100%	88%	100%	100%	100%
M1	100%	100%	78%	100%	
M2	100%	100%	86%	100%	100%
M3	100%	100%	100%	100%	100%
M4	100%	60%	73%	70%	89%
L1	100%	100%	100%	100%	100%
L2	100%	78%	100%	70%	80%
L3	100%	100%	86%	89%	93%
K1	100%	100%	100%	83%	40%
M5	100%	100%	100%	93%	100%
F	100%	100%	100%	100%	100%
F2	63%	71%	100%	33%	100%
E	100%	100%	100%		
O	100%	43%	57%	100%	100%
J	100%	86%	100%	83%	100%
Median	100%	100%	100%	91%	100%
Mean	97%	90%	90%	87%	93%
STD	9%	17%	14%	17%	15%
Grade 2					
Median	100	78	55	88	
Mean	100	74	62	84	
STD	0	23	19	13	

Table 2. Overview of % correct for Grade 1 (space reasons); Average results are given for both grades. Grade 2 was not recorded for most players during the last game.

RESULTS - GAMEPLAY

Each of the play sessions is recorded with a screen recorder.[4] The recording is then transcribed according to the example given in Table 3.

For each trial/visit and each step in the game, the list of words the child has to choose from is transcribed. Additionally, we record which word was chosen and whether this was correct. In addition each step time is recorded. As a result, we are able to compute the % correct, the number of words a child sees in the game, and how long each decision takes. Since interviews interrupted the timing in the fourth session, these times are excluded.

Results for each of the children are recorded in Table regarding their performance in terms of %correct choices taken, no clear trend is visible as can be seen in the example listing for Grade 1 in Table 2. What is remarkable however, is that while they often answer correctly, the main difference between the players seems to be their increase in speed. Figure 4 shows example timelines for a child from first and second grade each across all four sessions.

While the time between moves does not vary significantly across the two grades in the study, there were other differences. Separating children in each grade according to their performance on the written test, some significant differences in speed could be found. *N* Children are selected who fit best into one of each of these three groups: children with large changes in pre- and post-test ($N = 3$), those who remain weak ($N = 2$), and those who were strong from the start ($N = 2$). Since the first game ensures that the game mechanics are known for the second trial and the fourth trial was interrupted with interview questions, we compare time until next move for second and third trials only. The resulting distributions of speed for each group are shown in Figure 5.

By conventional criteria, the difference in time-play for the group with largest performance change in their written tests is considered to be very statistically significant, where N=21 moves (2nd trial), and 17 moves (3rd trial) with a p-value of 0.0015. There is no change in distribution of response time for the low performing children with no improvement in the final test. Even for those children who were already good in their orthographic abilities, the change in speed is significant. The two-tailed P value equals 0.0451, considered to be statistically significant, where N=12 moves (2nd trial), and 13 moves (3rd trial). For Grade 2, only the change in speed of the top performers is statistically significant (p=0.0231). Speed relates to automaticity, which in turn should facilitate actively using the skill in other tasks. This relationship might be visible in the data.

KNOWLEDGE TRANSFER

Theory On Writing

The analysis of lexical patterns in language acquisition are considered to be a complex procedure [2, 22, 12]. Orthography acquisition can be considered a form of language acquisition from oral to written representation [9]. In general, learning of patterns in complex processes goes through various phases of acquisition. These are studied in a large number of publications, especially in relation to dyslexia or aphasia patients.

The declarative/procedural model argues that lexicon and language depend on two neural systems that are intensively studied in the context of memory: declarative and procedural memory [23]. We believe that the change in speed is a form of moving from the application of a pattern from the procedural model into automatizing the cognitive procedure, similarly to how a driver commits the steps of changing gears from procedure to automation. Since the words presented to the children do not repeat very often, declarative learning (memorization that can be very fast) should not be taking place. Instead, the child should be forming a mental pattern that requires repetition with improved performance and speed rather than memorizing a set of words for a mental lexicon [14, 13]. The learning phase is highly individual as well as age related and depends on their previous knowledge. Compared to a first-grader, a third grader that learns a pattern will respond differently and might need more time to un-learn certain wrong ideas that have formed about orthography in the absence of teaching the pattern formally [1]. Not all children learned in the same way or with the same speed, but at the end of the four meetings, each child will have seen between 100-150 words and played an average of 6 minutes over the course of the month.

[4] AZ Screen Recorder for Android

Visit Number	Name	Date		Trial	Answer	Time	A	B	C	D	E	Correct / Inc	Emotion Det	Chosen	Missed	Distractor
1	O		Grade 2	1	A	00:27	miese	schönsten	röcheln			Correct	Normal	ie		
				1	B	00:33	schwärzer	miezte	Kratern			Correct	Normal	ie		
				1	D	00:40	schönsten	Fahrern	Knöcheln	Tiere	Gläsern	Correct	Normal	ie		
				1	B	00:45	Brutzeln	Kieseln	Hölzern	Schrubber	röcheln	Correct	Normal	ie		
				1	B	00:52	Klöstern	vieles	Kickern	schönsten	röcheln	Correct	Normal	ie		
				1	D	00:58	Schwärme	schwärzer	Löffeln	Schlieren	Knöcheln	Correct	Normal	ie		
2	O		Grade 2	1	B	00:25	brannten	Quieken	drinnen			Correct	Normal	ie		
				1	D	01:29	brannten	brüllten	stilles	hörten	kritzeln	Correct	Normal	?		tz
				1	E	02:06	könntest	miesen	spritzen	Dribbeln	Quallen	Incorrect	Normal	KV	ie	tz
				1	B	02:24	könntest	miesen	spritzen	Dribbeln		Correct	Normal	ie		tz
				1	C	02:43	stellten	Mannen	Fluren	pressten	brüllten	Correct	Normal	LV		
				1	B	02:48	stammten	Spiegeln	spritzte	Dribbeln	Flaggen	Correct	Normal	ie		tz
				1	D	02:59	solltest	stillen	konntest	Siedler	spritzen	Correct	Normal	ie		tz
3	O		Grade 2	1	A	00:13	zielen	Kletten	stillen			Correct	Normal	ie		
				1	B	00:25	drinnen	Strafen	stillen			Correct	Normal	LV		
				1	D	00:35	Flaggen	Krippen	Gruppen	Ziege	Drittel	Correct	Normal	ie		
				1	C	00:45	Kletten	Grotten	Dose	Platten	spritzte	Correct	Normal	LV		tz
				1	C	00:55	Gruppen	Brunnen	Miezen	Brillen	Flüssen	Correct	Normal	ie		
				1	C	01:06	Klassen	Schäfer	Grotten	Knappen	Platten	Incorrect	Normal	KV	LV	
				2	B	01:09	Klassen	Schäfer		Knappen	Platten	Correct	Normal	LV		
				1	B	01:24	Krippen	Fladen	stillen	drinnen	spritzte	Correct	Normal	LV		tz
4	O	07.06.201	Grade 2	1	C	00:10	klettern	klappern	tiefer			Correct	Normal	ie		
				1	B	00:18	klettern	Tierchen	schwitzen			Correct	Normal	ie		tz
				1	C	00:38	glitzern	schnappen	Nudeln	klappern	schnellen	Correct	Normal	LV		tz
				1	D	00:51	Klammern	Schritten	schleppen	schließen	schnitzen	Correct	Normal	ie		tz
				1	D	01:04	schlappen	schöne	schwitzen	glitzern	schmissen	Incorrect	Normal	tz	LV	tz,tz
				2	C	01:22	schlappen	schöne	schwitzen		schmissen	Incorrect	Normal	tz	LV	tz
				3	B	01:24	schlappen	schöne			schmissen	Correct	Normal	LV		
				1	B	01:39	Schlitten	Röcken	knattern	schnellen	Schlüssel		Normal	error Dict		ck

Figure 3. Sample Transcription of Game-run

Figure 4. Timing for Games; Child F is from Grade 1, Child M from Grade 2.

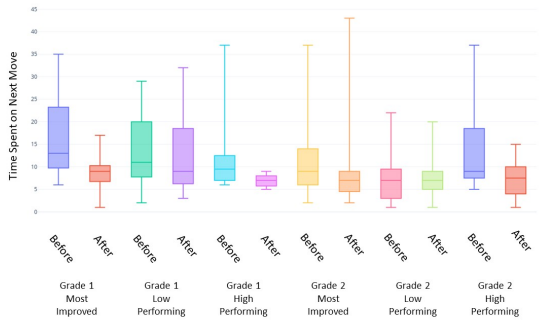

Figure 5. Changes in time spent on next move for Grade 1 and 2.

Theory On Reading

Understanding orthography should have an effect on eye-movement that can be measured with an eye-tracker [2, 16]. In addition, highly frequent words are no longer decoded but can be observed at the word level [25, 17]. This effects the amount of total fixation for particular words with different properties, including orthographic difficulty, length as well as frequency [11, 24]. This effect is most likely no longer observable on adults. Some experiments on adults for the same data confirm that the difference in reading for various orthographies and frequencies give us no distinguishable readings. Children, especially in the phase of reading acquisition, read differently and their decoding speed and probably confusions can still be measured [7, 8, 15, 10].

Correct reading ability is the precursor to reading practice, which in turn creates better readers (through self-teaching) and probably better writers (to be studied). Therefore, measuring progress in reading ability and training misconceptions is absolutely essential in the early phases of reading acquisition [20, 12]. We used a mobile eye-tracker that was mounted on a high-performance laptop to record children's readings of a text that was chosen based on a trial study with Grade 2. During this study, it became clear that measuring reading ability was more complicated than originally expected. We had to adjust the text and measurement features several times until we were able to narrow down what to look for and make sure this information was available in the text. As a result we are reporting the post-test for the eye-tracker in Grade 1 and only partially for Grade 2 as certain words were missing from the original text as we were still learning about that task of measuring reading ability.

It remains to be seen how skills acquired during game play might transfer to reading and writing abilities. This is the goal of the next sections.The present study makes no claims on cause and effect but reports the data that was collected. Given the findings however, it seems that there is a clear need to study this in more detail as there are indeed very strong indications that practicing with the games has an impact on reading ability of long and short vowels.

	pre		post			
	DC	ie	DC		ie	
Grade 1	33	6	47	(p=.11)	46**	(p<.0001)
Grade 2	56	90	71*	(p=.05)	79	(p=.88)

Table 3. Shown is % correct orthography in two areas trained in game, double consonant and correct usage of <ie> based on pre- and post-test results for each grade. (DC = double consonant)

		DC	ie
Grade 1	Improved	8	12
	Not improved	10	5
Grade 2	Improved	9	2
	Not Improved	9	3

Table 4. Number of children who increased their orthographic skill by more than 20% in each of the categories by grade, (DC = Double Consonant). The total number of children do not add up to the same number because some children wrote words that were not planned and therefore could not be measured on some feature.

MEASURING WRITING ABILITY

The pre- and post-test were transcribed from the tests according to what the child wanted to write vs. what they actually wrote. Most of the time, the word corresponded to the target that we had aimed for with the sentence or the picture. When this was not the case, the target consonant duplication and the <ie> target number of occurrences were changed accordingly. Next, the errors were counted and normalized by the amount of potential occurrences in their resulting list of words. Based on each child's own performance gain, three groups can be distinguished. Those that have improved, those that remain low performing and those that were already high performing in the pre-test. (These are the groups that we used to look at the speed change while playing.)

Table 3 reports the overall improvement in the results for both error categories. It can be seen that the largest impact was on <ie>, which was explicitly targeted on our first visit.[5] There was also a barely significant improvement in the double consonants for second graders.The teacher of this classroom started to introduce the concept to the children after our post-test was finished.

Rather than looking at the overall average of the classroom, Table 4 counts how many of the kids had significant changes of larger than 20% in each of the areas. In second grade almost no-one can improve on <ie> since the topic was mastered by the time of the pre-test. Regarding the double consonant, it can be said that about half the classroom has improved their score by at least 20%.

MEASURING READING ABILITY

When first starting to use the eye-tracker to find out whether there were differences between children in reading ability, we went through several rounds before focusing on a couple of features that seem to pin-point their ability very simply. The key to our break-through in measuring reading ability was firstly, to ask the children to read out loud so we could observe whether they were reading the words or skimming. At the same time, this slowed them down, which was good for the eye-tracker.

Our hypothesis was experience driven. Children learn that German orthography is shallow, which means they are only taught the long pronunciation of a vowel and they will sound out the letters as they are written: /mʊtɛʁ/ (Mutter=Muter-no meaning) and /rɔsɛ/ (Rosse=Rose-different semantic) and /wɪsɛn/ (Wissen=Wiesen-different semantic). Semantically mapping the word "Rosse" (horses) to "Rose" (rose) or "Wissen" to "Wiese" will cause problems for the reader and cause the reader to backtrack multiple times and finally guess which word would make sense and look similar. A sentence like "Wir Wiesen nicht wo die Rose lang laufen" ("We don't meadows where the rose are running" will take some time to decode. Since long and short vowels are quite frequent, struggling on every sentence for meaning, will not create readers.

We expect readers with no training on double consonants to be faster with L1 words than with L2 words since they have not understood the concept of the magic consonant.[6] However, if they understand double consonant concept (or have memorized the words), then L2, since it has a short vowel should be read faster. (This only works for child readers who are still decoding as they read. There would be no difference for experienced readers here.) In the first data collections, we found that it was hard to find any trends separating strong from weaker students. The breakthrough came, when we separated high-frequency (HF) words from low-frequency (LF) words. Low frequency words are unlikely to have been memorized by students. Therefore, they lend themselves well for separating the frequent readers (memorization) from those that understood the concept (generalize to new words). This concept resulted in four categories of words, namely L1-HF, L2-HF, L1-LF, and L2-LF. We found that the reading speed is comparable within these four word-groups for a given reader. In other words, picking out a specific word in a text could represent the skill for that group. Readers that were fast with the HF words, are expected to be frequent readers. If they were slow with L2-LF then they have not understood the pattern but did well with memorization on the L2-HF. The best readers would therefore show shorter reading time even for L2-LF because it shows they are able to generalize to unseen words, which improves reading (comprehension).

Figure 6 shows a prototypical distribution of the speed for all four categories for an example child from first grade in each of the three categories, top performer, low performer and most improved (according to the orthographic exams). It can be seen that those children from the "most improved" group are in the process of learning. The HF words are not

[5]To put these results into perspective: In our study another first grade went from 8% to 75% while the control-group in the same classroom went up to %50, another classroom went from %12 to %26.

[6]According to the Phontasia method, the easiest words to to read and write as they have bidirectionally the highest correspondence between phonemes and graphemes are 2-syllable words with long vowel [5]. These are called Level 1 (L1) words. Level 2 (L2) words add the smallest next-highest complexity to the orthography. They are 2-syllable words with double consonants (using the magic consonant), thereby making the preceding vowel short.

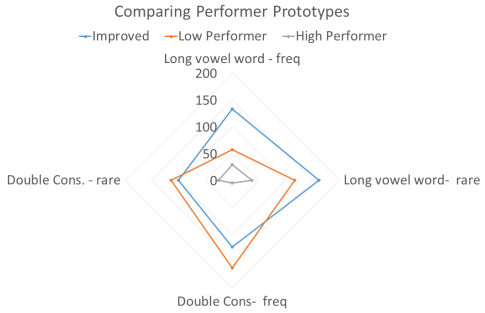

Figure 6. Examples of top, learning, and low performer, looking at total fixation duration.

automated yet, because the children have not had experience with frequent reading in first grade. So L2 and L1 words take about the same amount of time. This is the first step in the right direction. The top performers have an advantage with the HF words, perhaps because they have read more and started to memorize the words. They are able to read L2 faster than L1, for LF words this will be due to understanding the concept. This is what we aim for with the game practice. When we see this pattern, then it seems that the concept was internalized and can be applied to reading skill. A child that has not understood the concept of the magic consonant and has little reading experience to compensate with memorization can be recognized for the large amount of time spent puzzling over the meaning of an L2 word, whether it is HF or LF. In the graph only one example from each group is depicted for clarity reasons but the pattern repeats across children in each group.

Unfortunately, we do not have a pre-test for either group before the game play. We were only able to extract words from three of the four categories from Grade 2.[7]

Figure 7 plots the distribution over all children for each of the four words chosen to represent the four categories. L1-HF: "leben" (to live), L1-LF: "jagen" (to hunt), L2-HF: "können" (to be able to), L2-LF: "Welle" (wave), and L1-LF: "Feder" (feather) was used instead of "jagen" from the other text read by Grade 2.[8]

Both eye-tracker visits took place after the fourth visit of the study. In the diagram, we can see some overall tendencies looking at the total fixation duration. The first graders are comparable to the second graders. We are missing the L2-LF unfortunately, but in general, HF is faster than LF. The first graders performance for L1 on HF and LF are almost the same but in L2, they still have more difficulties parsing the LF words fast.

In future, it will be necessary to test the reading skill beforehand. However, seeing that the first graders are as strong as the

[7]We had to change the text to ensure that words from each of the four categories appeared. The original text was meant to be easy so it specifically did not contain rare words.

[8]Multiple trials with word choices for these categories have shown that they are pretty much easily exchangeable within category and exhibit the same reading behavior.

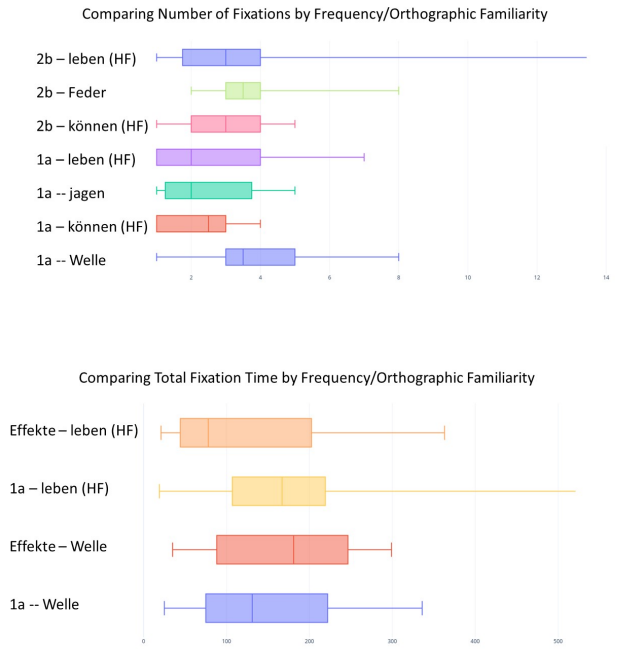

Figure 7. Overview of distributions for total fixation duration for each word of four categories.

second graders seems unusual and may or may not be the effect of the game training and should be verified across schools and classrooms. This data was compared to a collection from another event (EFFEKTE) where we tracked 5 children from first grade. For the Effekte group L1-HF (leben) was faster than L2-LF (Welle), indicating the struggle with the double consonant on unknown words. In contrast the Grade 1 from our study was slightly faster on L2-LF (Welle) than on L1-HF (leben), which is what we want to see. These are tendencies and not statistically significant and can be coincidental. Further studies are needed with more data. While these were explorative studies, the results give us a clear direction on how to set up a good study next year.

CONCLUSION AND FUTURE WORK

This paper reports on some results for a study that was performed when going into schools with games that are designed to practice and not teach concepts regarding the German vowel duration. We were able to look at aspects of game play such as accuracy and speed. The transfer of children's performance on orthographic and reading skills was studied. The learnings and the results are very useful for re-designing parts of the study for next year. Our impression from working across 10 further classrooms with a similar approach including Grades 1-3, the combination of teaching and practice with these games can be very powerful and the choice of grade when to apply this particular content seems to have a huge impact on the speed of learning. Due to the immediate feedback of misconceptions in the games as well as the large amount of automatically selected unique words shown to the children, the games reenforce pattern building over memorization. This seems to

help some children to generalize to low frequency words both in spelling and reading. While we only spent 6 minutes across four weeks with the children, we could already show tendencies of positive impact and we can only imagine what might happen when regular use of the game enters daily classroom routine. This will be our task to show over the final year of the iRead project.

REFERENCES

1. Hélène Beaunieux, Valérie Hubert, Thomas Witkowski, Anne-Lise Pitel, Sandrine Rossi, Jean-Marie Danion, Béatrice Desgranges, and Francis Eustache. 2006. Which processes are involved in cognitive procedural learning? *Memory (Hove, England)* 14 (08 2006), 521–39. DOI: http://dx.doi.org/10.1080/09658210500477766
2. M. Behrmann and D. Bub. 1992. Surface dyslexia and dysgraphia: dual routes, single lexicon. *Cognitive Neuropsychology* 9, 3 (1992), 209–251. DOI: http://dx.doi.org/10.1080/02643299208252059
3. Kay Berkling. 2017. Phontasia: A phonics game for German and its effect on orthographic skills-first corpus explorations. *WOCCI* 2017 (2017), 6th.
4. Kay Berkling and Rémi Lavalley. 2018. Automatic orthographic error tagging and classification for German texts. *Computer Speech & Language* 52 (2018), 56–78.
5. Kay Berkling and Uwe Reichel. 2016a. Wortstruktur, Orthographie und Didaktik: Die Relevanz der Vokallänge. In *System, Norm und Gebrauch – drei Seiten einer Medaille? Orthographische Kompetenz und Performanz im Spannungsfeld zwischen System, Norm und Empirie*, B. Mesch and Ch. Noack (Eds.). Schneider Hohengehren, Baltmannsweiler, 200–228.
6. Kay Berkling and Uwe D Reichel. 2016b. Progression in Materials for Learning to Read and Write - a Cross-Language and Cross-Century Comparison of Readers. In *Workshop on Child Computer Interaction*. 1–9. DOI:http://dx.doi.org/10.21437/WOCCI.2016-1
7. Hazel Blythe and Holly Joseph. 2011. *Childrenś Eye Movements during Reading*. DOI:http://dx.doi.org/10.1093/oxfordhb/9780199539789.013.0036
8. Aline Frey. 2016. Titre: Eye movements in children during reading: a review. *International Symposium for Educational Literacy (SILE/ISEL)* (12 2016).
9. Lila R Gleitman and Paul Rozin. 1977. The structure and acquisition of reading I: Relations between orthographies and the structure of language. *Toward a psychology of reading* (1977), 1–54.
10. Lynn Huestegge, Ralph Radach, Daniel Corbic, and Sujata M Huestegge. 2009. Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision research* 49, 24 (2009), 2948–2959.
11. Holly SSL Joseph, Kate Nation, and Simon P Liversedge. 2013. Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review* 42, 2 (2013), 207.
12. Christian Klicpera and Alfred Schabmann. 1993. Do German-speaking children have a chance to overcome reading and spelling difficulties? A longitudinal survey from the second until the eighth grade. *European Journal of Psychology of Education* 8, 3 (1993), 307–323.
13. Leonard F Koziol and Deborah Ely Budding. 2009. *Subcortical structures and cognition: Implications for neuropsychological assessment*. Springer Science & Business Media.
14. Leonard F. Koziol and Deborah Ely Budding. 2012. *Procedural Learning*. Springer US, Boston, MA, 2694–2696. DOI: http://dx.doi.org/10.1007/978-1-4419-1428-6_670
15. Brett Miller and Carol O'Donnell. 2013. Opening a window into reading development: Eye movements' role within a broader literacy research framework. *School psychology review* 42, 2 (2013), 123.
16. Anne Katrin Rau, Kristina Moll, Margaret J Snowling, and Karin Landerl. 2015. Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of experimental child psychology* 130 (2015), 92–105.
17. Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124 3 (1998), 372–422.
18. Martin Schöfl. 2016. Delfino–Ein Online-Förderwerkzeug bei LRS für den Einsatz in der Klasse. *Lernen und Lernstörungen* (2016).
19. Sascha Schroeder, Kay-Michael Würzner, Julian Heister, Alexander Geyken, and Reinhold Kliegl. 2015. childLex: a lexical database of German read by children. *Behavior Research Methods* 47 (2015), 1085–1094.
20. David L Share. 2008. Orthographic learning, phonological recoding, and self-teaching. In *Advances in child development and behavior*. Vol. 36. Elsevier, 31–82.
21. Yao-Ting Sung, Kuo-En Chang, and Tzu-Chien Liu. 2016. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education* 94 (2016), 252–275.
22. G. Thorstad. 1991. The effect of orthography on the acquisition of literacy skills. *British Journal of Psychology* 82, 4 (1991), 527–537. DOI: http://dx.doi.org/10.1111/j.2044-8295.1991.tb02418.x
23. Michael T Ullman. 2001. The declarative/procedural model of lexicon and grammar. *Journal of psycholinguistic research* 30, 1 (2001), 37–69.
24. Christian Vorstius, Ralph Radach, and Christopher J Lonigan. 2014. Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition* 22, 3-4 (2014), 458–485.
25. Stephen J White. 2008. Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of experimental psychology. Human perception and performance* 34 1 (2008), 205–23.