Automatic assessment and immediate feedback in first grade mathematics

Einari Kurvinen, Rolf Lindén, Teemu Rajala, Erkki Kaila, Mikko-Jussi Laakso,

Tapio Salakoski

Department of Information Technology
University of Turku
20014 Turku, Finland

{emakur, rolind, temira, ertaka, milaak, sala}@utu.fi

ABSTRACT

In this paper we present a study where automatic assessment and immediate feedback was utilized to support the learning of mathematical concepts and automatization of basic arithmetic calculations. Two first grade classes from a Finnish elementary school participated in the research. One class acted as a treatment group (N=23) and the other as a control group (N=21). Pupils in the treatment group worked for ten weeks one lesson a week with a collaborative education tool called ViLLE. Additionally, the pupils in the group did all of their homework with the system. Pupils in the control group used traditional learning methods. Before the experiment began a pre-test was conducted for both groups. The same test with three additional exercises was given after the ten week period. The pre- and post-tests included exercises to evaluate math skills and a questionnaire about pupils' attitudes towards math. The treatment group got statistically significantly better results compared to the control group in the post-test. In attitudes, the only statistically significant difference between the groups was in a question of how well they liked math. However, the difference in attitudes inside the groups was not statistically significant.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**]: Computer Uses in Education – *Compute-assisted instruction, Collaborative learning.*

General Terms

Design, Experimentation, Human Factors

Keywords

Computer-assisted learning, ICT in schools, Automatic assessment, Immediate feedback

Koli Calling '14,November 20 - 23 2014, Koli, Finland

http://dx.doi.org/10.1145/2674683.2674685

1 INTRODUCTION

Finland performed extremely well in Pisa (Programme for International Student Assessment) research in 2003, especially in mathematics. The success lifted Finland's educational system into worldwide awareness. However, in Pisa 2012 the learning results on math declined. [16, 21]. Another international study, TIMSS (Trends in International Mathematics and Science Study) support the findings of Pisa research. In 2011, the learning results of seventh graders in Finland dropped to the same level as the sixth graders in 1999 [20]. It seems that the traditional methods are lacking and there is a need for new ideas to promote learning.

This paper presents a study where computer-assisted learning (CAL) was utilized in first grade mathematics. The goal was to find out how well automatic assessment and immediate feedback can enhance learning. Math assignments were prepared and provided in a collaborative education tool called ViLLE developed at the University of Turku, Finland. ViLLE has extensive support for various types of automatically assessed exercises with immediate feedback. Additionally, before the study, new exercise types were implemented to ViLLE. The new exercise types were designed on the one hand to teach mathematical concepts to pupils and on the other hand to motivate them to spend more time in solving arithmetic problems.

In the first grade pupils will learn basic arithmetic calculations and it is important for future math skills that basic arithmetic calculations will became automatic [8]. Basic arithmetic calculations or arithmetic facts mean addition with one digit numbers and subtraction with numbers from 0 to 19 [1, 2, 6]. Ikäheimo has specified prerequisites which are needed in order to learn more advanced math skills later in primary school. Prerequisites for first grade are: numerosity, equality, number line back and forth, arithmetic facts, base-10 system and the concept of multiplication. [9]. Prerequisites by Ikäheimo are promoted by the Finnish Ministry of Education and are thus suitable in this research setting. Exercise types in ViLLE try to address these prerequisites in order to promote learning and learning results. According to Gersten et al. learning difficulties in math are unstable, which means they can be overcome with the right kind of support [8]. Visualization and automatic assessment is used to give this kind of support in ViLLE exercises.

Two usual limitations of exercise books can be overcome with computer-assisted learning. Firstly, a limitless amount of calculations generated for the pupils to solve. Secondly, immediate feedback and visualizations can used to support learning. Most of the exercises in ViLLE also make it possible for

pupils to choose one of the three difficulty levels. This helps keeping pupils motivated because they can work on their own skill-level.

This paper first discusses computer-assisted learning in other studies and the already known positive effects of it. After that, ViLLE is introduced in more detail, covering the basic ideology behind the design of the learning environment. Research problems and, methodology are discussed after ViLLE introduction, covering the used exercises, procedures and participants more profoundly. Finally there's the results and discussion about the main findings in this research.

2 RELATED WORK

Computer-assisted learning is not a new thing and it has been studied a lot. There are clear indications that CAL motivates students and improves learning results also in other fields than mathematics [13, 15, 26]. The studies do not however explain why CAL motivates pupils or why using CAL seems to enhance learning results. Many of the studies that will be discussed later in this section have incomplete research design. It is also very probable that intervention research has an effect if the intervention is well designed, no matter what equipment is used. This study concentrates on learning mathematics in elementary instruction, which means teaching in pre-school, first grade and second grade (ages 6–8 in Finland).

Computer assisted problem solving has been studied with 132 fifth graders in Taiwan. Pupils were divided into two groups, one control group and one treatment group. The treatment group practiced with MathCAL and the control group used traditional pen and paper methods. The research lasted four weeks, two times a week for 40 minutes. MathCAL utilizes Mayer's principles [19], where effort is put in problem solving and visualizing relevant information, instead of just the answer when solving math problems. Results from the study shows that CAL supported problem solving and helped to decrease pupils' cognitive load while working on problems. [4].

In another research Wilson et al. studied a game called Number Race to improve pupils' number skills. The goal of the gamer is to strengthen skills in number comparison and the connection between number word, number symbol and amount, using drilling. The study had only 22 7–10 year-old participants without a control group. Playing Number Race improved recognition of amounts, size comparison and subtraction skills. However no change was seen in addition or in understanding the base-10 system. [27].

The Number Race game was studied again by Räsänen et al. together with a game called Graphogame-Math. 350 pre-schoolers took part in the experiment. Only 30 of the participating pupils were chosen in the treatment group. All of the selected pupils had low numeracy skills. Rest of the pupils acted as a passive control group. All pupils in the control group had normal math skills. The Number Race -game concentrates in comparison of numbers and their relations whereas Graphogame-Math handles exact numbers. According to the research team, Graphogame-Math advanced reliability of number comparison the most in participating pupils. Furthermore, the Number Race -game showed similar results compared to the passive control group. However there is a lot of deviation in the results, which indicates that playing games is suitable for some pupils and not for others. Adequate amount of repetition is mentioned as a reason for the positive results by the research team. [23].

Also Fuchs et al. (2006) have studied CAL with pupils with math disability and reading disability. 16 pupils attended three times a week in math session, each lasting for 10 minutes. The goal of the sessions was to automate arithmetic skills with numbers 1–10. In pre- and post-test comparison pupils seemed to have improved their addition skills but no remarkable difference was seen in subtraction. This is the opposite what was seen in Number Race research by Wilson et al. [27]. There was also no transfer to word problems.

3 VILLE

ViLLE is a collaborative education platform developed at the University of Turku. ViLLE is web based so it's available from anywhere and is device independent. The exercises, whether they are automatically or manually assessed, are in very important role in development of ViLLE. Another important aspect is the ability to reduce teachers' work load. ViLLE enables teachers easily to create, share and asses courses, exercises and materials. Many exercise types in ViLLE have support for automatic assessment and immediate feedback.

ViLLE saves the student's progress scores and achievements. Collected information is shown to the student at all times to make learning with ViLLE more motivating. For teachers, the collected information is a valuable asset in assessing students' performance. Teachers can assess also the learning progress instead of just the end result. This also enables teachers to monitor the quality of teaching throughout learning period.

Compared to traditional math exercise books, ViLLE has three clear advantages, which are also examined in this research. One key advantage is endless amount of generated exercises. Most of the math exercises are generated by pre-set limitations. Every time pupil starts an exercise, new calculations are generated. Second advantage is automatic assessment. With ViLLE, pupils will see the correctness of their calculation immediately after solving the calculation. In traditional school work pupils have to check their answers by themselves after finishing one or more exercise book page. The third advantage is visualization of the problem and answer. With visualization mathematical concepts can be easily explained and used to support learning [5]. In math exercises different visualization can be used to make exercises more varying. Money, dots, number lines, ten blocks and written form can be used to visualize numbers which should help in solving problems. Figure 1 shows an example of automatic assessment and immediate feedback used in ViLLE exercises. The animated number line in the figure is also a good example of interactive visualization used in feedback.

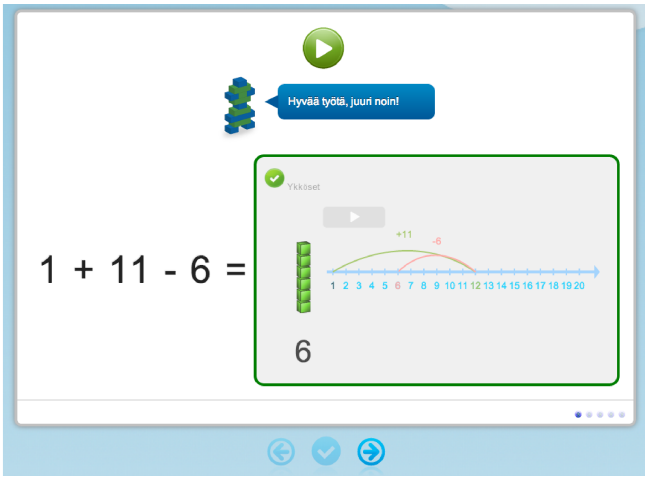

Figure 1 An example of automatic assessment and immediate feedback in ViLLE math exercises.

ViLLE has a wide variety of math exercises suitable to primary school. More exercise types were created for this research. These exercises include game-like drilling exercises like ladder-game, flying calculations, falling balls and three minute test. All these games introduce ether time limit, game like graphics or both to provide more motivating exercises to pupils.

There are several studies about ViLLE and its effect on learning performance. Most of the studies concentrate on university or high school students. The studies focus on learning performance on computer programming and collaboration. [10, 11, 12, 18, 22]. One study has previously been conducted on primary school mathematics. In the study third graders' learning performance on understanding decimal numbers and fractions were examined using pre- and post-test and a short one lesson intervention between the tests. The study showed some positive impact and encouraging results on both pupil motivation and learning result. [17].

4 RESEARCH

The experiment conducted in two first grades in Finnish elementary school evaluates the learning results of first grade mathematics with and without using ViLLE. The study is a typical pre-/post-test design between two groups and between subject factor is the usage of ViLLE.

The experiment was conducted in the beginning of school year which is the autumn term in Finland. The first graders had just started their school couple of weeks before the experiment began. The treatment group had one math lesson converted into computer assisted ViLLE lesson on each week during the ten week experiment. The control group proceeded with regular math curriculum approximately the same pace as the treatment group.

4.1 Research problems

There has been previous research utilizing ViLLE on studying learning results on elementary school mathematics. This research continues from those results with a more complete research setup and long term effects of CAL.

What is the effect of utilizing ViLLE on the learning results and attitudes towards mathematics in the first grade?

The effects of computer assisted learning were studied during the autumn term of the first grade in Finnish elementary school. The effectiveness of the ViLLE, and more generally CAL, is evaluated comparing the learning results of pupils in two different groups. In addition to learning results, also the attitudes towards mathematics are evaluated during the research.

4.2 Participants

Two first grade classes from a Finnish elementary school took part in the research. All parents gave permission for their children to take part in the research totaling 44 pupils, who took part in this research. One of the classes formed the treatment group (N=23) and the other class acted as the control group (N=21). One student from the control group was absent during the pre-test. The whole control group took part in the post test but two pupils were absent from the treatment group totaling 21 participants from the treatment group.

The treatment group integrated six pupils who got remedial instruction in contrast to the control group, which had none. Both of the classes participating utilized Varga-Neményi method (Hungarian mathematics) learning mathematics.

4.3 Materials

The pre- and post-test used in this experiment were specifically developed for the purpose of this study. Both the pre- and the post-test were two-parted. The first part measured learning results in math and the second part measured attitudes towards mathematics. Pre- and post-tests were identical, apart three additional exercises in the first part measuring learning results. The additional exercises were added to the test to make sure that there was room for improvement with the best performing pupils.

The first part of the pre- and post-test, which measures learning results, is based on the work of [14]. They present example exercises adapted from various math performance tests targeted for first graders. The example exercises were further adapted to be used in the first part of the tests. The exercises types in the tests are targeted to the end of first grade, hence some of the exercises are supposed to be demanding for the pupils.

The first page of the test consists of exercises which test pupil's knowledge on numerical symbols and numerals. In the first exercise pupils were supposed to circle the image, which consists of the asked amount of balls. Then in the second exercise pupils were requested to circle a certain numerical symbol. Third exercise measured counting skills of the pupils by asking how many lollipops there are in the picture. Each exercise was worth one point. Only in one case half point was given to a pupil, who counted lollipops on each row and not in total.

Exercises in the second page were more demanding. In the fourth exercise pupils were asked to continue number sequence (for example 2, 4, 6,…). The exercise consisted of four number sequences. Each sequence had three slots to put a number in. Each correct answer in a slot was worth half points. The idea behind this exercise was to see how well pupils could move back and forth in a number line using different intervals. Some pupils wrote two digit numbers mirrored (for example twelve as 21). If the answer was otherwise correct, these answers were interpreted as correct. This was due the fact that the pupils hadn't had formal education on two digit numbers yet. The fifth exercise measured

the understanding of magnitude of numerical symbols in numbers between 0–20. Pupils were required to circle the larger number out of two options. There were seven pairs and each correct answer added half points to total score. The sixth exercise tested pupil's addition and subtraction skills and the understanding of symbols used in the calculations. The exercise had three calculations which went past the tens. Altogether the sixth exercise had ten calculations and each was worth one point.

These six exercises formed the first part of the pre-test. Total score of the test was 22.5 points. In addition to these exercises the post-test included three additional exercises. The seventh exercise was similar to the sixth, except in seventh exercise the answer to the calculation was given and one of the numbers in the equation was hidden. The exercise measures arithmetic skills, deduction and more profound understanding of the number line. Two calculations out of ten went past ten. Each correct answer was worth one point. The eighth and ninth exercises were word problems. Equation was not required to get full scores. The eighth and ninth exercises were each worth one point. The extra page's total score was 12, adding the post-test maximum to 34.5.

The second part of the tests measured pupils' attitudes towards mathematics. The test was a short questionnaire consisting five questions. Pupils answered to the questions by circling smiley faces corresponding to statements. A happy face indicated agreeing with the statement and a sad face disagreeing with the statement. The smiley faces were consistent with Likert scale from one to five. Thereby the minimum score from the second part was 5 and maximum score 25.

Three of the five questions measured pupil's self evaluation of his/her skills in mathematics. One question measured how the pupil liked mathematics and one measured how useful they feel mathematics is. The questions are kept short because many of the pupils have just learned to read or are still learning.

The attitude test was pretested with one first grader, who could read. The pupil was interviewed during the test to make sure he had understood the questions properly. Only the last questions about usefulness of mathematics was difficult to understand but the researcher decided to keep the question regardless the difficulties.

4.4 Procedure

The research was carried out in the beginning of school year in autumn 2013 in a Finnish elementary school. Two first grade classes participated in the experiment. One class acted as treatment group (N=23) and the other as control group (N=21). The research started with a pre-test, which was then followed by a ten week long treatment period. Before Christmas break, the experiment was concluded with a post-test. Figure 2 describes the procedure used in this experiment.

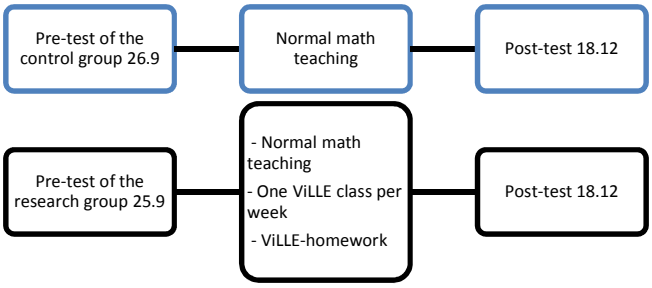

Figure 2 Procedure of the experiment

The experiment was introduced to parents and research permissions were gathered before starting the experiment. ViLLE and ViLLE's requirements were introduced to parents to make ViLLE based homework possible during the experiment.

The pre-test of the treatment group was conducted by the researcher while the pre-test of the control group was carried out by the teacher of the class. All the questions were read out loud because not everyone knew how to read yet. Especially fourth and fifth exercise needed more attention from the teacher or researcher. It took about 30 minutes to complete the pre-test.

Both groups followed the same curriculum, expect in the treatment group one mathematics lesson each week was replaced with a ViLLE lesson over period of ten weeks. These lessons were held on Wednesdays on split-lessons, when half of the class is present at once. First group consisted of ten pupils and the other consisted of thirteen pupils. Each lessons lasted for 45 minutes. There were no extra lessons compared to the control group. The control group had a normal "pen and paper" -lesson while the treatment group practiced math with ViLLE. For the first six weeks there were separate exercises to work on at school and home. For the last four weeks pupils were given larger amount of exercises but the same exercises applied also for homework. The more a pupil achieved at school, the less he had to work at home. This had a positive impact on motivation to work during the lessons and it also helped some pupils who encountered some technical difficulties at home. The pupils of treatment group had no other homework from math than ViLLE exercises. In the control group pupils had they regular homework form the math exercise book. During the first five weeks some pupils had to work in pairs due to insufficient amount of computers. The school managed to get enough computers for individual work from the sixth week of the experiment.

Time for a lesson is 45 minutes in Finland. During the first week of the study, the computers were prepared for pupils to conduct ViLLE exercises. From the second week onwards, pupils were required to practice some general ICT-skills like typing, using mouse, turning computer on, logging in and shutting down the computer. Average time to work with ViLLE exercise was about 15–20 minutes, which was enough to keep the pupils' engaged. New homework was given to pupils after each week's ViLLE-lessons. The parents were instructed to allow maximum of 15 minutes at a time for the homework.

The post-test was conducted by the researcher in both groups before the Christmas break. School assistant was present during the test to help children to read the instructions. Because the same test (apart from the three additional exercises) was used in the pre-test, pupils were instructed before the test and then they were allowed to advance in their own pace. In the treatment group everybody finished the test in 40 minutes in comparison to the control group, where two pupils had to be asked to stop working on the test after 45 minutes.

5 RESULTS

The results from the pre- and post-test were analyzed using statistical methods. The learning results are evaluated between the two groups. Attitudes towards mathematics are compared between the groups and between pre- and post-test in the group.

What is the effect of utilizing ViLLE on the learning results and attitudes towards mathematics in the first grade?

The experiment began with a pre-test for the both first grades taking part in the experiment. 23 pupils from the treatment group and 20 pupils from the control group took the pre-test. The tests were assessed by the scoring described in chapter 4.3. Table 1 summarizes the results from pre-test for the treatment and control groups.

Table 1 Statistical descriptive of the pre-test of treatment and control group.

	Treatment group	Control group
N	23	20
Mean	15.08	15.75
Median	16.50	16.00
Std. dev.	5.11	2.98
Min.	6	10.5
Max.	22.5	21.5

Even though the mean and median values are close to each other in groups compared, they are very different by composition. Treatment group's standard deviation is greater than in the control group, which means that there is more polarization in math skills in treatment group. Same trend can be seen in minimum and maximum scores. In the treatment group two students achieved the maximum score of 22.5. Also the lowest performing pupil in treatment group got 4.5 points less than the corresponding pupil in control group.

The normal distribution of these results was confirmed with Shapiro-Wilkin test (treatment group $p=0.143$ and control group $p=0.675$; $p>0.05$). Thereby the difference in averages between the groups was analyzed with independent samples T-test. The difference was not statistically significant ($p=0.601$; t-value: -0.528; df: 36.171; Levene's test: $F=11.276$; $p=0.002$), hence the groups were comparable in the beginning of the test. Table 2 presents the comparison between the groups on each exercise in the pre-test.

Table 2 Exercise comparison between groups

Exercise	Treatment (mean/std. dev.)	Control (mean/std. dev.)
1	1.00 (0.00)	1.00 (0.00)
2	1.00 (0.00)	1.00 (0.00)
3	0.96 (0.21)	1.00 (0.00)
4	2.17 (2.31)	2.68 (1.84)
5	3.39 (0.81)	3.48 (0.11)
6	6.57 (3.03)	6.60 (1.98)
Total	15.08 (5.11)	15.75 (2.98)

As can be seen in the Table 2, the control group achieved better mean in each exercise compared to treatment group. The corresponding results after post-test can be seen in Table 5.

Two pupils from the treatment group were absent from the post-test, totaling 21 from the treatment group and 21 pupils from the control group. Post-test results are analyzed in two phases. The first comparison is made with comparable parts of the pre- and post-tests. These results can be seen in Table 3. Second comparison includes also the additional exercises in the post-test, see Table 4.

Table 3 Statistical descriptives of post-test's identical part

	Treatment group	Control group
N	21	21
Mean	19.50	18.40
Median	21	19
Std. dev.	3.36	2.24
Min.	10	12
Max.	22.5	21.5

Compared to the pre-test, the post-test has greater difference in average scores. Also the median is higher in treatment group than in control group. Standard deviation is still higher in treatment group, although it has decreased. The difference in minimum score is decreased from 4.5 to 2 point. Also no one from the control group got maximum scores from the post-test. The normality of these results were analyzed with Shapiro-Wilkin, which showed that the results deviate from normal distribution (treatment group $p=0.019$, control group $p=0.01$; $p<0.05$).

Table 4 Statistical descriptives of post-test as a whole

	Treatment group	Control group
N	21	21
Mean	28.98	26.74
Median	32	27.50
Std. dev.	6.30	3.95
Min.	14.5	17
Max.	34.5	31.5

When comparing the descriptives from Table 4 to Table 3 we can see the same trends in mean and median, but with greater differences. Also the standard deviation is increased in treatment group. Table 5 shows the post-test results between the groups by exercises. Last three exercises are the additional exercises.

Table 5 Exercise comparison between groups in post-test

Exercise	Treatment (mean/std. dev.)	Control (mean/std. dev.)
1	1.00 (0.00)	1.00 (0.00)
2	1.00 (0.00)	0.95 (0.22)
3	1.00 (0.00)	0.88 (0.3.19)
4	4.07 (2.18)	3.85 (1.44)
5	3.26 (0.82)	3.40 (0.45)
6	9.17 (1.11)	8.35 (1.04)
7	7.90 (2.95)	7.35 (1.73)
8	0.81 (0.40)	0.40 (0.500)
9	0.76 (0.44)	0.60 (0.500)
Total	28.98 (6,30)	26.74 (3.95)

When comparing the results of Table 5 and Table 2, the treatment group achieved better mean in every exercise, except in exercise 5, which was the size comparison of the numbers.

The normality of treatment and control groups was also determined with Shapiro-Wilkin which shows that the data deviates from normal distribution also in this case (treatment group $p=0.014$, control group $p=0.01$; $p<0.05$). Thereby the results from post-tests are analyzed with non-parametric Mann-Whitney U-test. The results are shown in Table 6.

Table 6 Post-test comparison between the two groups

	Pre-test	Post-test (comparable)	Post-test
Treatment group's mean	15.08	19.50	28.98
Treatment group's std. dev.	5.109	3.36	6.30
Control group's mean	15.75	18.40	26.74
Control group's std. dev.	2.98	2.24	3.95
p		0.023	0.008

In both comparisons the treatment group performed statistically significantly better than the control group. When comparing the whole test the result is statistically very significant. The learning results were also clear ($p<0.00$) on both groups, when analyzed with Wilcoxon Signed Rank -test. The usage of automatic assessment combined with immediate feedback in ViLLE has helped the treatment group to perform better than the control group.

The second part of the test measured pupils' attitudes towards mathematics. The attitude test measured four different aspects of attitudes: liking mathematics, self evaluation of math skills, usefulness of mathematics and total scores in the test. These results can be seen in Table 7, Table 8, Table 9 and Table 10. The self evaluation on math skills was collected from three different questions, which were compiled to a sum variable. Reliability of the sum variable was confirmed with Cronbach's alpha (0.856, >0.6).

The reliability of the formed sum variable was tested with Cronbach's alpha in each test for each group. All the values are over 0.6, hence the sum variable can be held reliable.

Table 7 Treatment group's results on attitudes in pre-test

	Like	Usefulness	Skills	Sum
N	23	23	23	23
Mean	3.52	4.61	12.74	20.87
Median	4	5	13	21
Std. dev.	1.44	0.66	2.16	3.07
Min.	1	3	9	15
Max.	5	5	15	25

Table 8 Treatment group's results on attitudes in post-test

	Like	Usefulness	Skills	Sum
N	21	21	21	21
Mean	3.33	4.62	12.19	20.14
Median	3	5	13	21
Std. dev.	1.32	0.74	2.73	3.75
Min.	1	3	5	11
Max.	5	5	15	25

All the four measured variables show positive feedback from the students. Liking mathematics and the usefulness is measured on scale from 1 to 5. Skills are measured on scale from 0 to 15 and the overall result is thereby from 0 to 25. The median shows even more clearly that the pupils have a positive attitude towards mathematics and are confident about their math skills. They also feel that mathematics is useful for them. When comparing the pre- and post-test, the difference in answers were not statistically significant.

Table 9 Control group's results on attitudes in pre-tests

	Like	Usefulness	Skills	Sum
N	20	20	20	20
Mean	3.85	3.75	12.00	19.60
Median	4	4	13	20.50
Std. dev.	1.18	1.68	2.71	4.77
Min.	1	1	3	5
Max.	5	5	15	25

Table 10 Control group's attitudes in post-test

	Like	Usefulness	Skills	Sum
N	21	21	21	21
Mean	4.10	4.48	12.81	21.38
Median	5	5	13	22
Std. dev.	1.18	0.93	2.205	3.74
Min.	1	2	6	9
Max.	5	5	15	25

The results from the control group are also very positive. The averages and medians all increased from pre-test to post-test, even though the only statistically significant change is in usefulness. The pupils feel more confident on their math skills, like math more and feel that it's more useful according to the pos-test. The deviation of the answers has decreased, which means that the pupils agree more on these points.

The treatment group thought math to be more useful in the post-test questionnaire. The deviation of the answers is also smaller in treatment group than in the control group, apart from the questions about liking math. Only the liking of mathematics was better in control group than in the treatment group. The changes in answers after the treatment are however the opposite between the groups. In the control group all the averages went up, some even past the treatment group. Even though the deviation decreased in the control group, they are on about the same level in the post-test than the treatment group.

The normality assumption was abandoned based on skewness and curtosis of the answers. The comparison between measurements inside the group was analyzed with non-parametric Wilcoxon Signed Rank -test. The only statistically significant development was in usefulness of mathematics in control group $p=0.032$ ($p<0.05$). It should be noted though that the treatment group feels that math is more useful after the post-test even the value decreased.

The results between the groups were analyzed using independent samples U-test. The control group seemed to like mathematics more after the post-test than the treatment group ($p=0.046$; $p<0.05$). The change in liking between measurements comparing individual groups was however not statistically significant.

6 DISCUSSION

Based on the results of this study ViLLE has a positive impact on learn basic arithmetics. The treatment group learned mathematics better than the control group. The finding was also statistically significant ($p=0.023/0.008$). The difference in average is significant event though the sample size is relatively small and nonparametric U-test was used to analyze the result. In the pre-test the control group achieved better or same average on each exercise than treatment group. In the post-test result was other way around. The treatment group achieved better or same average on each exercise, except on exercise 5. The scale in the post-test was not enough for the treatment group. Two students got the maximum score, while in the control group best performing student got three points less than the maximum score.

What makes these findings even sturdier, is that two pupils from the treatment group were absent during post-test. These pupils were tested a month later, including two weeks of Christmas holiday, than the original treatment group. These test subjects scored 11.5 and 10.5 points from the pre-test, which is well below the average score in the group. In the delayed post-test these pupils scored 34.5 and 31 points, which are well above the average of the group. These two pupils seemed to benefit a lot from doing ViLLE-exercises. However one could argue that these pupils got some extra training because of the delayed test. Even if this is the case, these pupils still managed to improve their results from one of the lowest to one of the highest scores. Also the pre- and post-test measured math skills that should be learned and mastered after first school year, hence the one month period doesn't give that much advantage.

Improvement of the results can be explained due the fact that treatment group calculated approximately 250–350 calculations per week compared to regular exercise book which has approximately 20-40 calculations per double page. First graders have four math lessons per week which totals 80–160 calculations plus homework, which is usually 10–20 calculations. In school pupils used 5–20 minutes, depending on pupil, to do ViLLE-exercises during the lessons. At home pupils spent 5–60 minutes working on homework per week. Almost without an exception one or two pupils failed to do any homework or only finished one or two exercises. About half of the pupils managed to achieve decent scores to show effort and engagement. Some pupils had difficulties with exercises with audio or more game like features, which naturally affected their scores and time spent doing exercises. Even though the parents were instructed to divide ViLLE homework to 15 minutes slots, about one third of the pupils did all the homework during the last night before next ViLLE-lesson.

ViLLE's math exercises have a built in feature to support different challenge levels. This enables well performing pupils to choose more challenging exercises and low performing pupils to choose more basic exercises. This helps in motivating pupils and lets them to learn at their own level.

Besides doing more work, ViLLE-exercises are developed to support learning by automatic assessment and immediate feedback. These features support automatization of basic arithmetic facts or calculations [24, 25]. Learning of mathematic concepts is promoted by visualization of calculations and numbers in various exercise types.

When examining pupils' attitudes towards mathematics, it should be noted that all the answers are very positive and well above the average. The drop in liking mathematics between pre- and post test is not in line with previous studies. Multiple studies show that computer assisted learning increases motivation of the pupils [3, 4, 13, 15, 26, 27]. Even though the liking of mathematics did not change statistically significantly in repeated measures analysis, the overall result, control group liking mathematics more than the treatment group, was statistically significant. Many of these studies have been criticized being short and the motivation being caused by new teaching methods. This experiment lasted 10 weeks, which is long enough to exclude this effect. One explanation for the decreased liking could be the increased amount of work the pupils did during the experiment. The treatment group also solved a lot of different kinds of exercise types, some being more challenging than an average math exercise. This means that some exercises were harder and might have affected the attitudes towards mathematics. Besides these points, also the other mathematics lessons have their input on these results. Nevertheless even though the liking decreased, pupils stayed motivated and seemed to enjoy doing ViLLE-exercises.

Some pupils encountered minor technical difficulties when working at home or at school. At school the wireless base station got overloaded from time to time which caused some delays and frustration. Also the Windows computers used in this experiment could update and reboot in the middle of the lesson. These technical problems might have affected attitudes of the treatment group.

Also the feeling of usefulness of mathematics decreased in treatment group and increased in control group. It should be noted though that even with this trend the feeling of usefulness stayed higher in treatment group than in control group.

Based on anecdotal evidence, pupils in treatment group learned also general computer skills in addition to math skills. These skills

include handling a laptop, starting computer, logging in to windows, launching browser, navigating in browser and shutting down computer properly. Pupils also had to take care of their own login credentials to ViLLE.

7 CONCULSION AND FUTURE WORK

ViLLE, the use of automatic assessment and visualization had a clear positive impact on learning performance of the pupils in treatment group. Attitudes towards math, usefulness of math and self evaluation of the math skills decreased little in treatment group during the experiment. Only the liking of mathematics was statistically significantly different between the two groups in the post-test. To make any profound conclusions on attitudes would need a bigger sample and classes from multiple schools.

The results on learning performance should also be verified duplicating this research setting in larger scale including multiple schools and classes. Also the time spent on experiment could be made longer to examine study paths in different performance groups. More profound research project has already begun together with city of Salo, which aims to create complete set of electronic exercises for classes 1–6. This project also enables studying learning performance among multiple classes. Delayed test could be used to track long term effects of computer assisted learning.

8 REFERENCES

[1] Ashcraft, M. H. 1992. Cognitive arithmetic: A review of data and theory. *Cognition, 44* (1), 75–10

[2] Baroody, A. & Rosu, L. 2006. Adaptive expertise with basic addition and subtraction combinations: The number sense view Baroody AJ, Torbeyns T. chairs. Developing Adaptive Expertise in Elementary School Arithmetic. Symposium conducted at: The annual meeting of American Educational Research Assosication.

[3] Beauchamp, G. & Kennewell, S. 2008, The Influence of ICT on the interactivity of teaching. *Education of Information Technology 13* (4), 305–315

[4] Chang, K. E., Sung, Y. T., & Lin, S. F. (2006). Computer-assisted learning for mathematical problem solving. *Computers & Education, 46* (2), 140-151.

[5] Cheng, P. C. H. 1999. Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education 33* (2), 109–130.

[6] Domahs, F. & Delazer, M. 2005. Some assumptions and facts about arithmetic facts. *Psychology Science 47* (1), 96

[7] Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. 2006. The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39* (5), 467–475.

[8] Gersten, R., Jordan, N. C., & Flojo, J. R. 2005. Early dentification and interventions for students with mathematics difficulties. *Journal of learning disabilities, 38* (4), 293–304 .

[9] Ikäheimo, H. 2011. Solmukohdat. Edu.fi. Read 13.1.2014 from http://www.edu.fi/erityinen_tuki/materiaaleja_oppimisen_tueksi/matematiikka/s

[10] Kaila, E., Rajala, T., Laakso, M.-J. & Salakoski, T. 2011. Important features in program visualization. Appeared in ICEE: An International Conference on Engineering Education, 21–26 August 2011, Belfast, Northern Ireland, UK.

[11] Kaila, E., Rajala, T., Laakso, M.-J. & Salakoski, T. 2010. Long- term Effects of Program Visualization. In 12th Australasian Computing Education Conference (ACE 2010), January 18–22, 2010. Brisbane, Australia.

[12] Kaila, E., Rajala, T., Laakso, M.-J. & Salakoski, T. 2009. Effects, Experiences and Feedback from Studies of a Program Visualization Tool.Informatics in Education, 8, 1, 17–34.

[13] Kebritchi, M., Hirumi, A., & Bai, H. 2010. The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & education, 55* (2), 427–443.

[14] Koponen & Aunio. 2008. Lukimat. Esimerkkitehtäviä ensiluokkalaisten arviointiin. Haettu 26.8.2013 osoitteesta http://www.lukimat.fi/matematiikka/materiaalit/tulostettava-materiaali/Arviointi_1luokka.pdf

[15] Kulik, C. & Kulik, J.A. 1991. Effectiveness of Computer-Based Instruction: An Updated Analysis. Julkaisussa: *Computers in Human Behavior, 7*, 75–94.

[16] Kupari, P., Välijärvi, J., Andersson, L., Arffman, I., Nissinen, K., Puhakka, E., & Vettenranta, J. (2013). PISA12 ensituloksia: Opetus-ja kulttuuriministeriön julkaisuja 2013: 20.

[17] Kurvinen, E., Lindén, R., Rajala, T., Kaila, E., Laakso, M.-J. & Salakoski, T. 2012. *Computer-assisted Learning in Primary School Mathematics Using ViLLE Education Tool.* 12th Koli Calling International Conference on Computing Education Research, November 15th to 18th, 2012, Tahko, Finland.

[18] Laakso, M.-J. 2010. Promoting Programming Learning. Engagement, Automatic Assessment with Immediate Feedback in Visualizations. TUCS Dissertations no 131

[19] Mayer, R. E. 1992. Cognition and instruction: Their historic meeting within educational psychology. Journal of Education *Psychology, 84* (4), 405–412.

[20] Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). TIMSS 2011 International Results in Mathematics. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

[21] Pisa 2012: Suomalaisnuorten osaaminen laskussa. 3.12.2013. Opetus- ja kulttuuriministeriö. Fetched 25.5.2014 from: http://www.minedu.fi/OPM/Tiedotteet/2013/12/pisa.html

[22] Rajala, T., Laakso, M.-J., Kaila, E. and Salakoski, T. 2008. Effectiveness of Program Visualization: A Case Study with the ViLLE Tool. In: *Journal of Information Technology Education: Innovations in Practice, 7*, IIP 15–32.

[23] Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer- assisted intervention for children with low numeracy skills. *Cognitive Development, 24* (4), 450–472.

[24] Siegler, R. S. 1988. Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development 59* (4), 833–851.

[25] Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. 2009. Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education, 24* (3), 335–359.

[26] Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K. & Wright, M. 2006. Computer gaming and interactive simulations for learning: a meta-analysis. *Journal of Educational Computing Research, 34* (3), 229–243.

[27] Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. 2006. An open trial assessment of" The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions, 2* (1), 20–36.