Earth, Life & Social Sciences
Kampweg 5
3769 DE Soesterberg
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

www.tno.nl

T +31 88 866 15 00
F +31 34 635 39 77

TNO report

TNO 2016 R11141

E-learning Sudan
Final report Phase II

Date	August 2016
Author(s)	H.E. Stubbé M. van der Klauw J.J. Langefeld N.C.M. Theunissen A.H. van der Hulst
No. of copies	4
Number of pages	115 (incl. appendices)
Number of appendices	8
Sponsor	War Child Holland
Project name	Conn@ct.Now / e-Learning Sudan
Project number	055.01039

Management Summary

After a successful first pilot with a mathematics game developed to support autonomous learning in remote areas in Sudan, a second, more substantial pilot was carried out. The aim of this second pilot was to confirm the learning outcomes of Pilot I, in a larger pilot: more children and a longer period of time. At the same time, this pilot intended to investigate and provide more information on psychosocial effects of the introduction of digital, autonomous learning in remote villages. Finally, in this pilot game interaction research was included, to better understand what children think of the game, and which elements they like best.

This report provides the theoretical background for the mathematics game. Based on this theory, an optimal set of game requirements is described. Then the game design is explained, showing how the game requirements are integrated in the design. The remainder of the report describes Pilot II and its results. The conclusions, relating to the research questions, are summarised below.

Do children learn mathematics by playing the game for a longer period of time (sustained learning)?
This pilot has shown that all children improved their mathematical abilities by playing the game. The fact that there is a correlation between total amount of time (months) played and the post-test scores also shows that the children continue to learn when they play the game for a longer period of time. There were no significant differences for gender. The differences between age groups, with older children performing (slightly) better than younger children, is to be expected and follows the normal development of children. There are significant differences between states, with White Nile performing best on test A, and Gedaref having the lowest scores on test B. Children in North Kordofan show the largest increase in scores between pre and post-test. There are significant positive correlations between the three methods used to measure learning effects (mathematics tests, logged data, and EGMA). This means that all three methods show similar results, which makes the overall conclusions about learning effect stronger. These findings also indicate that in the future it is not necessary to use all three methods, when the game is used in the same or a similar context.

What are the (psychosocial) effects of learning with technology on children and the communities they live in?
To answer this research question a questionnaire, focus group discussions and interviews were used. Unfortunately, only three constructs of the questionnaire were reliable: Self-esteem, Self-efficacy, and Motivation and future orientation. Self-esteem has significantly increased during the pilot. Self-efficacy and Motivation and future orientation have not changed significantly. This means children have got a better opinion about themselves, either because they have learned mathematics or because of other factors in relation to the pilot: e.g. social aspects of learning together, visits to the community by others, or the use of ICT. In addition, both Self-esteem and Self-efficacy show a significant positive effect on learning outcomes. The focus group discussions and interviews show positive effects as well.

Comments made during the focus group discussions included: *'The children like to learn, they like the game, have started playing together more, show better behavior, and keep their hands cleaner. Parents appreciated the fact that they knew their children were well taken care of during the learning sessions. Now children have started learning, they would like to continue, and learn more subjects as well.'*

What do children think of the game?
All the children liked the game. They thought it was fun and liked the colours.
One child said it had a bit of difficulty. Unfortunately, only two constructs of the questionnaire about game interaction were reliable: Usability and Game improves level of knowledge. The average scores on these constructs were (slightly) below average, but showed no significant correlations with learning results. The qualitative evaluation showed that both boys and girls liked the game, but they appreciated different elements (mini-games and visions). This may explain why boys and girls show similar learning results; although their preferences may differ, they could find elements in the game they liked.
The scores on: 'The mini-game works', and 'I know what to do in this mini-game' were average with very little variation.

What factors contribute most to learning effects?
A multivariate regression analysis on all the factors that were reliable, and/or showed significant change during the pilot, showed that the scores on the pre-tests contributed most to learning effects. This does not mean that the knowledge children have before starting the game determines the results. Children with a lower score on the pre-test had a significantly higher increase of scores than children with a higher score on the pre-test. Although ceiling effect and statistical regression might influence this, this shows that children learn from playing the game. This is supported by the fact that percentage complete (logged data) contributes to learning outcomes: the higher the percentage of the game completed, the higher the score on the post-tests. As mentioned before, Self-esteem and Self-efficacy have a positive effect on learning outcomes. The longer the distance to the nearest primary school, the higher the learning effect. Children might be more motivated if there is no real alternative to access education. The level of education of the mother and having a father both have a negative effect on learning outcomes. At the moment, it is not clear why. Gender and age do not relate to learning outcomes, which is a positive finding: boys as well as girls, from different age-groups, can learn from the game.

What are the characteristics of the children who dropped out?
As drop out is always an issue in pilots that run for a longer period of time, specific attention was paid to this. A total of 62 children dropped out, 10% of the children. According to the remarks made by facilitators, the reason for 5% of the drop out was that either parents of children 'refused the programme'. The other 95% of the drop out was, according to facilitator remarks, caused by families moving to another community to find water or harvest the crops. Analyses show no significant differences for gender or age. Boys as well as girls, of all age-groups, have dropped out. Although this is a positive finding, as it shows that the intervention does not exclude specific groups, it does not provide any information that could be used to decrease drop out. Children in White Nile had a significant higher risk of dropping out. The explanation for this is that in two communities in White Nile almost all children have dropped out. Facilitators reported that many families in these communities had moved during the pilot period.

In addition to the results, this pilot has also provided insight in a number of research challenges:

- Test protocol
  Although facilitators and observers had been trained to use the test protocol, tests were not (always) taken at the designated times.
- Data collection
  Collecting the data took much time, with researchers travelling to the communities to do the testing. Results were written down on paper, and later entered into an Excel file. This was time-consuming, and allowed for human error at the different stages of data entry.
- Control group
  Using a control group is also an issue: there is an ethical element in asking communities to participate in a pilot as a control group without allowing them to benefit from it. In addition, agreements have to signed at various levels before communities can participate. As a result the control groups were smaller than the experimental group.
- Unique child numbers
  Some of the facilitators used the children's names for the accounts on the tablets, in English and in Arabic. Due to this, only 508 of the 532 log files could be matched to children.
- Logged data
  As there was no reliable internet connection, logged data had to be uploaded manually and brought to Khartoum by hand. Due to the time involved, logged data were only collected at the end of the pilot and could not be used to feed back into the programme during the pilot.
- Timestamp
  As the tablets were not online, there was a problem with the timestamp for 20% of the tablets. The number of months played, as registered in the management system, was therefore not correct, and could not be used for further analysis.

Contents

Appendices

1 Introduction

Education for children in the developing world is in crisis. Children growing up in complex emergencies are at the sharp end of global development challenges: of the 58 million out-of-school children globally, 36% live in countries scarred by war and violence. More than one third of refugee children globally are missing out on primary education (UNICEF, 2015) and the safety and education of girls are disproportionately affected (Jones & Naylor, 2014). Reaching those without access to school is a pressing issue, particularly girls and children in the rural areas in Africa and South Asia (Burnett, 1996; Kallaway, 2001; Cremin & Nakabugo, 2012). Complex and mutually reinforcing patterns of disadvantage – poverty, gender inequity, disability, conflict and displacement – raise barriers to schooling and erode educational opportunities for children.

In recent years, global policy frameworks on education reflect the '*shift in the global conversation on education from a focus on access to access plus learning'* (UNESCO, 2013) because access to education does not guarantee that children will learn once they have access. The proposed Sustainable Development Goal (SDG) on education along with the Incheon Declaration adopted at the World Education Forum in 2015, outline a vision for sustainable and equitable education spanning pre-primary through to at least lower secondary. There is a recognition of the need to focus on quality as well as on access, and to support the most disadvantaged children, not least those living in emergencies. There are not enough qualified teachers, or support for teacher professional development; class sizes are large and under-resourced; and classroom methods are teacher-centred. This leads to high drop-out levels, up to 50% (Sriprakash, 2010).

Like many countries in the global South, Sudan struggles with the same issues. School attendance in the rural areas is about 50%, and is as low as 0% for some communities in some States. At present, it is not realistic to believe that this will be solved through traditional means. Any effort to make traditional basic education accessible for today's Sudanese children would require substantial investment, and the current government financial allocation to education is insufficient.
New solutions are urgently required for the current generation of out of school children living in Sudan: solutions which tackle issues of access, equity and quality. Innovative education technology, when coupled with a strong partnership and context specific approach, could provide these children with a previously unimagined chance to access quality education opportunities.

Whilst it has been generally accepted that educational technology by itself cannot improve learning outcomes (Power, Gater, Grant & Winters, 2014: Daniel, 2010), an agreement has not been reached on the most effective ways to utilise it.
There are many examples of unsuccessful e-Learning programmes in Africa (Bitew, 2008), but there are also a few notable examples of open and distance learning for early literacy, basic and primary education, and as useful tools in bridging the gaps of access due to conflict, gender, and geography (UNICEF, 2009). Uptake of education can be improved by offering flexible learning opportunities, including through the use of technology which can take educational access to a child in their location. Thus, learning can take place alongside the demands of the family and without perceived risk to children who might otherwise have to travel long distances.

In addition, learning should lead to an official certification and/or allow children to progress into regular education (UNICEF, 2009).

Online and distance learning with ICT are seen as possible solutions. The focus of this approach should be on rural areas, communities affected by conflict – including Internally Displaced Persons (IDPs) - and specifically include girls and minority groups. Digital technologies, if the software is well-designed and the content grounded in a well-constructed curriculum, can deliver one-to-one interactive instruction, in a consistent manner, to all children. Children can learn at their own pace and repeat material as often as they need. The use of technology also allows for individual monitoring of progress.

Promoting understanding of mathematics in the early years is critical, as longitudinal research has shown that early mathematical understanding is highly influential on later mathematics and reading performance at school (Duncan, Claessens, Huston, Pagani, Engel, Sexton, Dowsett, Magnuson, Kubanov, Feinstein, Brooks-Gunn, Duckworth & Japel, 2007), even after controlling for other basic skills that are known to affect school performance. For this and other reasons, specific to the context, we chose to develop a mathematics game. Across countries children learn mathematics in a similar way and learning of mathematics depends less on language. Moreover, there are many successful examples of applications and educational games that support the learning of mathematics. Although these are mostly enrichment materials, used in addition to classroom teaching, they provide a strong basis for the development of a mathematics game that supports autonomous learning.

The curriculum in the game is based on the official curriculum for out-of-school children in Sudan for Grades 1, 2, and 3 and leads to an official certification (Stubbé, Badri, Telford, Oosterbeek & van der Hulst, in press). The game supports autonomous learning and consists of two game worlds and various mini-games to practise each mathematics concept. It provides instruction, and has a management system that tracks progress and ensures that children do the mini-games that match their knowledge and skills. In addition, language and graphics are designed to be culturally appropriate. By playing the game, children are actively involved and can learn at their own pace. A more detailed description of the requirements for the game and the game itself is provided in Chapter 3.

1.1 Pilot I

As this way of learning is completely different from the formal school system in Sudan, and as the targeted children in remote, rural communities have never been to school, a proof of concept was needed before the three-year curriculum could be developed into a game. The most important question that needed to be answered was if children can learn mathematics by playing the game. Following a research driven approach - the ’fastest route to failure’ -, a small part of the game (six weeks of the curriculum) was developed and tested in three communities (60 children) in the period November 2012 - January 2013. The control group consisted of 20 children in a fourth community who did not receive any education in the same period.

Results show a significant increase of children’s scores on an oral mathematics test, which doubled from 19.4/60 on the pre-test to 38.4/60 on the post-test (Stubbé, Badri, Telford & van der Hulst, 2015; Stubbé, Badri, Telford, van der Hulst & van Joolingen, in press). Children in the control group, a fourth community, did not increase their scores in the same period; their pre-test score was 16.5/60, their post-test score 17.2/60.

Having thus proved that children can learn mathematics by playing the game, the project aims to test if children can learn by playing the game for a longer period of time, learning various mathematical concepts, and generate a body of research that will help to understand the social impact, contribute to the knowledge on e-learning, and pave the way for scaling-up of the project.

1.2 Pilot II

Pilot II, the present study, aims to repeat the learning results of pilot I with a larger group of children (591), in more states (3) and more communities (19), over a longer period of time (six months). The control group consisted of 325 children in ten more remote communities enrolled in informal education in out-of-school centers. As in Pilot I demographic and geographic information was collected and educational research was carried out. In addition, psychosocial research and game interaction research was included to explore and establish the effects of learning and technology on the children and the communities in which they live.

The research questions in this pilot were:

1 Do children learn mathematics by playing the game for a longer period of time (sustained learning)?
  - 1.1 Are there any learning effects in general?
  - 1.2 Are there any differences in learning effects for gender, age or state?
  - 1.3 Are there any correlations between the three methods with which learning effect is measured?

2 What are the (psychosocial) effects of learning with technology on children and the communities they live in?
  - 2.1 What are the effects of learning with technology on factors that are known to influence learning and motivation to learn?
  - 2.2 Are there any other effects of learning with technology on the children and the communities they live in?

3 What do children think of the game?
  - 3.1 Do the children think the game works?
  - 3.2 Do the children understand what to do?
  - 3.3 Do children like playing the game? Which characteristics did they like best?

4 What factors contribute most to learning effects?
  Geographical factors (state, locality, village), demographic factors (age, gender), psychosocial factors (self-esteem, self-efficacy, …), interaction with the game or level of mathematical competence at the start of the pilot?

5 As drop out is always an issue in pilots that run for a longer period of time, specific attention was paid to this. How many children dropped out? What are the reasons for children to drop out? What are the characteristics of children who drop out?

1.3 This report

This report describes the mathematics game, its theoretical basis, the research method and the research results of pilot II. Chapter 2 describes the theoretical background of the development of mathematical skills, an overview of the literature on ICT and education - with a specific focus on educational gaming - and a summary of the latest insights in and success factors of education in emergencies. Chapter 3 summarizes the requirements of the game and describes the game-design.
In Chapter 4, the research method is described. Chapter 5 shows the results of Pilot II. In Chapter 6 these results are discussed. This chapter also includes conclusions and the answers to the research questions included in this introduction.

1.4 Partners

E-Learning Sudan is conceived through a collaboration between the Ministry of Education of Sudan, Ahfad University for Women in Khartoum and War Child Holland. It is funded through the Ministry of Foreign Affairs of The Netherlands, with additional funding from UNICEF Sudan. Curriculum, game development and research is provided by TNO. The game was produced by Flavour with support from creative partners in Sudan.

2 Theoretical background

In this chapter the theoretical background of the project is described concisely. Paragraph 2.1 focuses on the development of mathematical skills, Paragraph 2.2 describes the literature on ICT and education, with a specific focus on educational gaming, and Paragraph 2.3 summarizes the latest insights in and success factors of education in emergencies. At the end of each paragraph, game requirements based on the literature discussed are provided.

The development of mathematical skills, ICT and education, and education in emergencies are three different research domains. The game, meant to support the autonomous learning of mathematics in developing countries depends on all three of these domains. Consequently, an optimal combination of requirements for the game was drawn up. This optimal set of requirements may not always be the best choice for each of the research domains separately. In Paragraph 2.4 the dilemmas related to determining this optimal set of requirements are discussed.

2.1 Development of mathematical skills

Research shows that children develop mathematical skills before beginning formal schooling (USAID, 2009). Across cultures, children seem to bring similar types of skills to school, but do so at different levels (Guberman, 1996). In general, children from low-income backgrounds begin school with a more limited skill set than those from middle-income backgrounds. This is related to the environment in which children grow up; that enables them to understand the world, master language and get insight in the basic knowledge needed for mathematics (Greenman, Bodovski & Reed, 2011). This means that children with a more limited set of skills will need additional support to ensure success (Chard, Baker, Clarke, Jungjohan, Davis & Smolkowski, 2008). Where in developed countries this support is usually given as extra support to individual children, in developing countries early interventions should be aimed at all children.

2.1.1 Development rate

The rate of acquisition of mathematical skills can be influenced by the opportunities children have in their communities (Guberman 1996). Household tasks and chores can get in the way of developing these skills, but they can also enhance the acquisition of these skills because they provide meaningful learning opportunities. Once children begin formal education, they use this informal knowledge when completing new tasks (Baroody in Copely, 1999; Ginsberg & Russel, 1981).

Becoming efficient at mathematics requires the automatisation of the subsequent stage, rather than repeating the earlier stages. Children need to free up cognitive resources to be able to solve more complex problems (Pellegrino & Goldman, 1987). With continued practice, children become more confident in their computational and problem solving skills (Fuson in Kilpatrick, Martin & Schifter, 2003). This puts a significant emphasis on good early mathematics experiences for children.

2.1.2 *Mathematical knowledge in the early years*

Between the ages of 3 and 9, the construction of number knowledge develops in more or less the same ways for all children (USAID, 2009). With formal schooling, children start developing a new understanding of mathematics. With continued practice, they become more familiar with numbers and their values, and their confidence grows; children process information faster in solving mathematical problems.

Across countries, curricular and conceptual goals show similar subjects, in the following order (USAID, 2009):

(1) developing an understanding of whole numbers, including concepts of correspondence, counting, cardinality, and comparison;
(2) representing, comparing, and ordering whole numbers, and joining and separating sets;
(3) developing understandings of addition and subtraction, and strategies for basic addition and subtraction facts, including whole-number relationships (e.g. tens and ones); and
(4) developing understanding of base-ten numeration system and place-value concepts, including fluency with multi-digit addition and subtraction.

2.1.3 *Game requirements*

First of all, a mathematics game comprising the curriculum of the first three grades of primary education should include the subjects listed above. In addition, as this project focuses on vulnerable children with little learning support from parents or teachers, we assume them to have little informal knowledge. The opportunities to learn from everyday life situations in the remote communities in Sudan are scarce. Due to this, the approach for struggling learners is followed. One of the major issues in supporting struggling learners is to make sure that there is a strong basis to build on. This corresponds with the concept of mastery learning (Bloom, 1985), where 'the students are helped to master each learning unit before proceeding to a more advanced learning task'. Furthermore, struggling learners need explicit instruction (Timmermans, 2005; Milo, 2003). Research (Bodovski & Farkas, 2007) shows that struggling learners show less engagement during instruction. If this engagement is increased, performance increases as well. A focus on 'time on task' (Carroll, 1963) could help to improve learning results; all children can learn mathematics, but some need more time than others.

To support struggling learners, instruction and exercises on the mathematical skills that are often acquired informally are included in the game. In addition, direct instruction is given: instruction that explains how to 'do it'. The language of instruction is basic, formal Arabic, which most children understand well.
The instruction in the videos is provided by slightly older children (14-15 years), which is assumed to increase motivation to watch the instruction videos. Furthermore, these older children can also be seen as role models, increasing motivation to learn and self-efficacy.

2.2 ICT and education (educational games)

The introduction of educational technologies has not changed human beings' fundamental capacities to learn, but it has profoundly changed how ideas and practices are communicated (Beetham & Sharpe, 2007). It is arguable that there are really no models of e-learning per se – only re-enhancements of existing models of learning. The challenge is to describe how technology allows underlying processes common to all learning to function effectively (Mayes & de Freitas, 2007). The role of technology may be primarily to get remote learners into a position to learn as favourably as if they were school-based, rather than offering a new learning method. It is more a new model of educational delivery than a new model of learning. Clark, Yates, Early & Moulton (2008) argue that the choice of media does not influence learning. Differences in instructional design prevail over the method of delivery. Usually the difference between passive and active learning and the clear and concise instruction determine variation in learning outcomes.
This means that in the use of educational games the design of instruction delivered by these games, including gamification, is of crucial importance. This is in line with Wouters, van der Spek & van Oostendorp (2009) who argue that the alignment of learning outcomes and game type, the alignment of game complexity and human cognitive processes, attention for cognitive and motivational processes, research on specific mitigating effects, like gender, on game effectiveness should be considered in game design.

Various meta-reviews and meta-analyses have shown the cognitive and motivational effects of educational games in general. In their meta-analysis of 32 studies in which traditional classroom teaching was compared to computer gaming or interactive simulation, Vogel, Vogel, Cannon-Bowers, Bowers, Muse & Wright (2006) found an overall positive effect of educational games: significant higher cognitive gains and a more positive attitude towards learning were observed in subjects using interactive simulations or educational games versus traditional teaching methods. This seems to be the case for boys as well as girls, although the low number of studies that reported statistics for males and females gives reason to consider these results with caution. All age groups showed significant positive results for the use of computer gaming or interactive simulation. The type of activity did not appear to be influential; neither did realism of the pictures in the game. Wouters, van Nimwegen, van Oostendorp & van der Spek (2013) investigated whether educational games are more effective in terms of learning and more motivating than conventional instruction methods. In their meta-analysis of 39 studies they found that educational games were more effective in terms of learning and retention, but were not more motivating than conventional instruction methods. Learners in educational games learned more than those taught with conventional instruction methods when the game was supplemented with other instruction methods, when multiple training sessions were involved and when players worked in groups.

Research on digital mathematics interventions has shown increased motivation (Rosas, Nussbaum, Cumsille, Marianov, Correa, Flores, Grau, Lagos, López, López, Rodriguez, & Salinas, 2013), more positive attitudes towards mathematics (Ke, 2008), and a better mastery of mathematics (Praet & Desoetel, 2014; Steenbergen Hu & Cooper, 2013; Li & Ma., 2010; Räsänen, Salminen, Annio &

Dehaene, 2009) for children in kindergarten and primary education. All of these studies were conducted with European or North American children. Recently, Pitchford (2015) evaluated the effectiveness of a tablet intervention for mathematics in a school in Malawi. She concluded that tablet technology can effectively support early year mathematical skills in developing countries if the software is carefully designed to engage the child in the learning process and the content is grounded in a solid well-constructed curriculum appropriate for the child's development stage.

2.2.1 *Effective game characteristics*

The use of technology should enable **active learning**, with a focus on the activities and interaction of learners, instead of on content in the sense of pre-prepared learning materials (Collis & Moonen, 2001; Jonassen, Peck & Wilson, 1999).
Another predominant game design feature that determines whether educational games work is that of the **level of control** that learners have over pace, order and strategy (Garris, Ahlers & Driskell, 2002). It determines whether learners have a sense of 'agency' and actually feel that they can influence events in the game.
The more learner control, the higher learner motivation. When children learn autonomously by playing an educational game, the game itself should provide **feedback** on correctness of their answers. Where we traditionally place this role with the teacher, sometimes computer-programmes are even better: they are able to always, consequently give feedback, something which in under-resourced, overcrowded classrooms is unlikely. In addition, *adaptive feedback* should be given, i.e. feedback on process as well as on results. Children should be invited to react, think again and be given tips to find the right answer. This approach increases attention and thus time on task (Carrol, 1963), which in turn positively influences learning results.
While learning without the support of a teacher, children will most likely not stay **motivated** over a long period of time unless they are intrinsically motivated. Learners that are intrinsically motivated engage in the learning process for ' … its own sake, for the enjoyment it provides, the learning it permits, or the feelings of accomplishment it evokes.' (Lepper & Iyengar, 1999:349). Also, the more intrinsic the motivation, the more durable the learning may be (Trinder, 2013).
In the tradition of Vygotsky (1978), learning is largely seen as a **social activity**. Learning should not simply be the assimilation of new knowledge; it is the process by which learners are integrated into a knowledge community. As learning is essentially a social phenomenon, learners are likely to be motivated by rewards provided by the knowledge community (Cremin & Arthur, 2014).

2.2.2 *Autonomous learning*

The most important constraint for the game design was that children will use the mathematics game in their own remote communities. This automatically implies that there are no school or teachers to explain the mathematical concepts. Although autonomous learning is often used for advanced learners and adults, it is not a commonly used approach for beginners. One of the most important issues that are highlighted in the literature with *self-learning* for beginning learners is that children do not have enough knowledge to follow instruction or do the exercises without further help - capacity to understand instruction (Vogel, Vogel, Cannon-Bowers, Bowers, Muse & Wright, 2006). This may be a result of the language of instruction: children do not understand the instruction. It can also be due to the instruction itself, which assumes a certain level of knowledge and cannot be understood by children who do not have this knowledge. Lastly, cultural appropriateness can influence the capacity to understand instruction: if the game is not culturally appropriate, children need to put in extra effort to understand the context; cognitive capacity that cannot be used to understand the mathematical concept.

2.2.3 *Game requirements*

The mathematics game aims to provide a rich learning experience based on an experiential learning approach: children should learn from active engagement with the environment (Dewey, 1938; Kolb, Boyatzis & Mainemelis, 2000). For learning to be effective, the children should be actively involved in many different mini-games. This increases the number of exercises they do and contributes to the time spent on learning. In this way, children can repeat and practise new concepts many times, without having the feeling that it is repetitive (Jonker & Wijers, 2008). For example, for children to master a new concept, they should have access to reproductive and productive exercises. The mini-games should provide several types of activities ranging from a ‘multiple choice’ activity where children choose the right answer, to ‘matching’ and ‘arranging’ numbers and amounts and finally ‘writing’ the correct answers to problems on the tablet.

To motivate children to actively use the mathematics game for a longer period of time, they should have a certain level of control over what they do first, the pace in which they play the game, and whether they want to go back to subjects they have completed already.The game should provide feedback on the correctness of children’s answers. In addition, children should be given the opportunity to receive hints and tips that help them to find the right answer.

The mathematics game should include elements that invoke intrinsic motivation to learn. There are three different possibilities for this: (1) The game itself should be appealing to the children: graphics, colours, and the narrative should be recognizable and fun. (2) The instruction in the game should be easy to understand, to the point, and easy to relate to. (3) Game-based learning is driven by extrinsic rewards and rewarding that stimulate intrinsic motivation. A crucial factor in intrinsic motivation in educational games is challenge (Malone, 1981) and competition. The game itself, or the setting in which it is played, should allow for some challenge and competition. Therefore, the game itself, or the setting in which it is played, should allow for social interaction between children, but also between children and adults. Adults should take up the more general classroom management tasks of teachers, like organizing the use of the tablets and helping with technical problems.

Children should be able to understand the language of instruction used in the game. The explanation of mathematical concepts given in the instruction should be easy to understand, never assuming that children have a certain level of knowledge and understanding. Any instruction and explanation should be culturally appropriate to reduce cognitive load.

2.3 Education in emergencies

Research (UNICEF, 2009; Bitew, 2008; Unwin, 2009) shows that educational programs with or without the use of ICT have not always been successful in developing countries. The three most important factors for an educational program to be successful are (UNICEF, 2009): (1) location, (2) flexibility, and (3) continuity. Specific needs identified were (1) an appropriate location for accessing learning materials and supplementary face-to-face contact, (2) flexibility in learning alongside other demands of the family, which might interrupt a traditional school schedule, and (3) the opportunity for progression into the mainstream educational system if desired by the learner and their family. This is in line with a more general conclusion of Clarke (2002) ‘in order for technology to improve learning, it must ‘fit’ into students’ lives … not the other way around’. The educational game and context in which it is played should reflect these success factors of educational programs in developing countries.

2.3.1 Game requirements

The game should support learning in the remote communities where the children live. The implication of this is that there are no teachers and schools. The game should, therefore, support autonomous learning.
The game should be flexible which means that it should support learning, also when children have to do household tasks. The game should, therefore, support self-paced learning, allowing children to follow an individual learning path; skipping a few days whenever household tasks prevent joining learning sessions, and continuing at their own level when they can.
The game should support educational continuity. Children should prepare for official exams in Sudan, which allows them to enrol in regular education. This means the curriculum in the game must be the same as the official curriculum for out-of-school children and be endorsed by the Ministry of Education in Sudan.

2.4 Dilemmas

The first three paragraphs of this chapter have provided an overview of the literature on the development of mathematical skills, ICT and learning, and education in emergencies. Each of these overviews leads to a number of specific requirements for the game. These requirements, however, differ per research domain. Sometimes they even contradict each other, leading to dilemmas when trying to decide on the optimal set of requirements for the game. These dilemmas are discussed in the following paragraphs.

2.4.1 Learner control versus guided learning

The first dilemma is the trade-off between learner control and guided learning. Although research shows it is more motivating to allow learners to have a certain level of control within the educational game (Garris, Ahlers & Driskell, 2002), the situation of illiterate children learning autonomously requires a much more guided learning environment. As learning has to take place without the instructional support of teachers, the approach for struggling learners was chosen, which actually leads to a more guided way of learning (direct instruction and a guided learning path on the basis of mastery). We have tried to partially solve this dilemma by creating a two-level game: the game world in which children have a certain level of control, and the mini-games in which learning is completely guided. If educational games are developed for more advanced learners – higher grades – we should look into possibilities to allow a higher level of control.

2.4.2 Anywhere, anytime

The second dilemma refers to access to the game. One of the most important advantages of game-based learning is that it can take place anywhere and anytime. This increases the possibility for children to learn. In the pilots in Sudan, tablets are shared. This means the tablets are locked up and only given to the children to use during learning sessions. The pragmatic reason behind this was cost-effectiveness: two children can use one tablet at different times of the day and the tablets are less likely to break. Another disadvantage of this approach is that children who move to another village during the pilot cannot bring the tablet and have to stop playing the game.

2.4.3 *Adaptive feedback*

The third dilemma is that it remains difficult to provide adaptive feedback. Research shows that two types of feedback have a positive effect on learning: feedback on the correctness of an answer and feedback regarding learning strategies. Although feedback with respect to a right or wrong answer is easy to integrate into an educational game, it is harder to provide hints and feedback on specific learning strategies. It is not always possible to determine what strategy children have used, by simply logging their answer. This makes it hard to suggest a better strategy. Some hints, especially regarding numbers, are included in the mathematics game. Nevertheless, more hints and feedback on learning strategies could be designed.

2.4.4 *Social interaction*

Finally, there is a dilemma regarding social interaction. Although research shows that children learn more and are more motivated by social interaction, the need to track individual progress children requires the children to use their own device, individually. This means children do not learn in pairs or groups. They are, however learning in learning sessions: more children learn at the same time, in the same place. During the learning sessions interaction can take place in the form of children helping each other or showing their progress to others.

3 Requirements and game design

In this chapter a more detailed description of the mathematics game is given. Then the relation between the game requirements and the game design is described. Finally, the similarities and differences between the game requirements and game design are discussed.

3.1 Iterative design

The ELS educational game for mathematics was developed in three phases. The first version allowed for six weeks of learning and was tested in a small pilot during December 2012 - February 2013. Based on the findings in this pilot, the second version was developed that allowed for six months of learning. This version was tested in a large pilot between October 2014 - March 2015. Based on the results of this large pilot, the final version of the game was delivered in the beginning of 2016, covering three years of learning.

3.2 Game design

The resulting mathematics game incorporates two distinct levels, each with a different pedagogy. The first level is that of Game Worlds which provide the connecting narratives for the second level, that of separate mini-games (44 different mini-games, 160 variations of mini-games). Game World 1 (see Figure 1) is about helping other children to achieve goals in their lives; by doing mini-games, children help other children to become e.g. a goat herder or doctor. Half of the jobs are familiar roles within the target communities, such as a cook, tractor owner or brick maker. The other half are known to the children, but belong to the outside world, like a teacher, nurse, doctor and engineer. In a playful way this helps the children to broaden their future perspective. Game World 2 (see Figure 2) is a shop where children can buy and sell products. By playing the mini-games, children can increase the number of products they can sell and enhance their shop.

Figure 1 Game World 1: building the community.

Figure 2 Game World 2: buying and selling in the shop.

The top level of the game environment, that of the Game Worlds, uses a predominantly *experiential learning approach.* Learners have a certain level of control over their exploration of the Game World. For example, they decide whether they watch an instruction video (when and how often they watch the videos without limits), check the progress they have made, do a mini-game or just try out the funny elements in the Game World. Within the shop narrative children can also decide themselves which products to buy and sell. The lower level, of the mini-games, has a different pedagogy, that of mastery learning with direct feedback on performance, and consequently less control over the environment. Each mini-game (see Figure 3) addresses a specific mathematical concept. Some mini-games have variations that can be used for several mathematical concepts and all mathematic concepts can be practised by several mini-games. This was done to help the children understand the mathematical concept and stay motivated at the same time. Progress through the mathematics game is based on performance: the number of correct answers within a certain time-frame decides whether children can continue to a more difficult mathematical concept. This ensures that children always work at their own level, at their own pace.

Figure 3 Screenshots of the mini-games.

To track individual progress of all children, the mini-games at the lower level are played individually. S*ocial interaction* elements are not included in the game. To compensate for this, the children play the game in learning sessions, supervised by a facilitator. Creating a situation in which different children learn at their own pace, but at the same time, is motivating and stimulates social interaction and competition where children can help other children or show their progress. Also, we expect them to talk about the development of their shop and about the jobs they have uncovered and thus reflect upon their learning. In addition, the narratives at the top level help to build intrinsic motivation; in a playful and implicit way, the game transmits the message to children that they can achieve more in life if they work hard.

3.3 Relation between game requirements and game design

Table 1, below, shows how the game requirements were used in the game design.

Table 1 Relation between game requirements and game design.

Game requirements		**Game design**
Subject	**Requirement**	
Mathematics curriculum	Include the mathematics goals of the official Sudanese curriculum for out-of-school children	In collaboration with the National Council for Literacy and Adult Education and the National Council for Mathematics the specific learning goals for mathematics of the official curriculum for out-of-school children in Sudan were included in the game.
	Include the conceptual mathematics goals that are similar across countries	A mapping of the Sudanese curriculum for out-of-school children in Sudan and the conceptual mathematics goals that are similar across countries, showed that some of these goals were not included in the Sudanese curriculum. In the game, the missing mathematics goals were added to the official Sudanese curriculum.
	The Ministry of Education should approve the curriculum in the game	The Ministry of Education in Sudan has approved the curriculum of the game.
Struggling learners	Use the mastery learning approach	The game guides children through a carefully designed learning path. They learn at their own pace, though: children can only move to a new/more difficult mathematical concept once they have mastered the previous one. This is measured by the number of mistakes made within a certain timeframe.
	Use direct instruction	The game contains 'instruction videos', short films, with children as actors, in which mathematical concepts are explained. Each instruction video explains one mathematical concept. Per mathematical concept only one strategy is explained to solve the problem.
	Heavy focus on 'time on task'	The game is attractive and motivating, which keeps children focused on it. The mini-games are an important part of the game, in which children actively engage and practise their mathematical skills.
Experiential learning	Active learning, with many, different activities	The mini-games within the game stimulate children to try out mathematical concepts themselves. All mathematical concepts can be practised with a variety of mini-games ranging from multiple choice, and matching answers to giving answers to problems by writing them down. The mini-games with which a mathematical concept can be practised always start with real-life situation (concrete), move to a representation of real life (model), and end with abstract problems.

Table 1 (resumed) Relation between game requirements and game design.

Subject	**Requirement**	**Game design**
Level of control	Learner control over what children do first	The game provides limited control over what children can do first. Children can decide to explore the game world, watch instruction videos, check what they have learned already and play mini-games – either new ones or repeat earlier ones. Children have no control over the order in which they play the new mini-games.
	Learner control over pace	The game allows children to play at their own pace, based on the number of correct answers within a certain timeframe. Children who only make a few mistakes progress through the game quickly. Children who make more mistakes will take longer to master the mathematical concept.
	Learner control over going back to previous activities	The game allows children to go back to previous activities whenever they want.
Adequate feedback	Feedback on correctness of answers	The game provides immediate feedback on the correctness of answers, using audio and visual elements.
	Feedback on strategy/process	The game does not provide specific feedback on the chosen strategy or process.
	Hints and tips	The game provides some tip and hints. E.g. at the beginning of the game children can click on a ‘voice-button’ that will count the objects in an answer.
Motivation	Game should be appealing	The graphics and the colours of the game are based on co-design with the children. The children made drawings of their own environment: clothes, food, animals, plants and trees. The graphics in the game are based on these drawings. The game worlds were designed after thorough discussions with local partners. The narratives in the game worlds are recognised by the children and should appeal to them.
	Instruction	The instruction videos in the game are short and attractive, and have children as actors. The children in the instruction videos acts as role models, thus motivating the children to learn as much they have.
	Extrinsic awards that stimulate intrinsic motivation	There are several motivating elements in the game: progress is shown with stars. When a child reaches a new level, a star is added to the progress bar. At the same time fireworks go off and there is applause. In Game world 1 the huts become more beautiful, in Game world 2 you can sell more objects in your shop.

Table 1 (resumed) Relation between game requirements and game design.

Subject	**Requirement**	**Game design**
Social interaction	Social interaction between children	Within the game it is not possible to have social interaction between children. But, children play the game in learning sessions: a group of children playing the game individually, at the same time in the same place. There is social interaction between the children during the learning session.
	Social interaction between children and adults	Within the game it is not possible to have social interaction between children and adults. But, children play the game in learning session: a group of children playing the game individually, at the same time in the same place, supervised by a facilitator. There is social interaction between the children and the facilitator during the learning session.
Capacity to understand instruction	Children should be able to understand the language of instruction	The language of instruction is a simple version of Modern Standard Arabic (the way you would speak to children).
	Instruction should be easy to understand, without the need for extra knowledge	The instruction videos explain all necessary mathematical concepts, starting from the very beginning (kiindergarten level). All instruction that builds on earlier instruction videos repeats the content of those videos before introducing the new, more complex concept.
	Culturally appropriate	All objects used in the game are objects that the children know in their everyday life. All the people are appropriately dressed and e.g. do not show the soles of their shoes. At the same time, much attention has been given to neutral examples (fruits and vegetables, bricks) to avoid references to conflict or minority groups. There are boys as well as girls in the game and as actors of the instruction videos.
Education in emergencies	Flexibility	The children play the game 45 minutes a day. This leaves time to add reading & writing to the learning sessions and still stay within two hours a day spent on learning. Children can skip a few days of learning, if their household tasks get in the way of joining the learning sessions, and continue at their own level when they come back.
	Locality	The game is played in remote villages. As there are no teachers in the villages, the game is designed as a self-paced, autonomous learning intervention.
	Continuity	The curriculum in the game prepares children for the official mid-primary school exam and thus leads to continuity: if they pass the exam children can proceed into regular, formal education.

3.4 Similarities and differences between game requirements and game design

Almost all requirements have been incorporated into the game design. The only exception being the four requirements that caused the dilemmas discussed in Chapter 2. (1) Children have some control over the game (pace, and going back to earlier materials), but they cannot decide what they want to do first. The reason for this is that a guided learning approach works better for struggling learners. Therefore, children will follow a specific learning path through the game, but can do so at their own pace. (2) The use of tablets and self-paced learning software enables children to learn ‘anytime and anywhere’. The advantage of this is that children can spend more time on learning, and show the tablet to their siblings and caregivers.
The disadvantage of giving tablets to children instead of having learning sessions is that the social interaction between children and between child and adult can disappear completely. Furthermore, tablets cannot be shared between children, which is less cost-effective. They are also more likely to break or get stolen.
The tablets are, therefore, only used in learning sessions. (3) The game does provide feedback on the correctness of the answers, but no feedback is given in relation to learning strategies. The most important reason for this is that it is quite difficult to do this digitally. (4) The last requirement that was not integrated in the game-design is social interaction. Although collaborative learning is quite effective and motivating, it makes it very hard to track individual progress. Tracking individual progress and supporting self-paced learning is more important in this context than the digitalised social elements. As the children learn in learning sessions, there is face-to-face social interaction.

4 Method

This pilot used a pre-test post-test control group quasi-experimental design. Participants were 591 children in 19 remote communities in three states in Sudan (White Nile, North Kordofan and Gedaref), aged 7-9. 51% of the children were girls, 49% were boys. 47% of the children were seven years old, 29% were 8 years old, and 24% were nine years old. The control group consisted of 325 children in 10 remote communities in the three states. The experimental group used the tablet game for a period of approximately six months, for a maximum of five days a week, 45 minutes a day, while supervised by a facilitator. The children in the control group were enrolled in informal education; they attended two mathematics lessons a day of 45 minutes each, taught by a teacher in out-of-school centers. In addition to the mathematics tests A and B, the internationally validated Early Grade Mathematics Assessment (EGMA) was taken by independent consultants, with a stratified sample of the children (210) in the experimental condition. Finally, logged tablet data of 532 learners were collected the tablets. Apart from educational research, geographical and demographic information was collected and psychosocial and game interaction research was carried out.

4.1 Set-up of Pilot II

4.1.1 Sample selection

In collaboration with the Ministry of Education in Sudan, communities were selected on the basis of geographical location: in all three states clusters of two or three communities within a short distance of each other were approached. In this way, travelling time could be reduced. When the communities agreed to participate, all children in the relevant age-group were invited to participate in the pilot.

4.1.2 Participation

Parents were informed about the goal and the method. The community was involved in setting up the 'learning centers' (huts where the children gathered to learn). Children were assigned to either morning or afternoon learning sessions, according to their parents' wishes. In this way, learning could fit in with their chores and household tasks. The communities could also select their own facilitator, choosing from three trained facilitators.

4.1.3 Learning Sessions and Hardware

As there were two learning sessions a day, the hardware could be shared. Each tablet was used by two different children. Consequently, hardware stayed in the learning center, locked away until the next session.

4.1.4 Facilitator

Each community had a facilitator. The facilitator was responsible for 'classroom management tasks' for example: encouraging the children to work with the mathematics game and helping with technical problems. The facilitator was not supposed to teach or explain the principles of mathematics. Facilitators were trained to take this role and to solve technical problems. During the week, they lived in the communities, in the weekends they could go home. The facilitator training consisted of a combination of face-to-face meetings and distance learning, including discussions and assignments. The curriculum included child-friendly approaches, educational background of the game, and technical aspects of the game and tablet.

In general, communities had two facilitators, taking turns supervising the learning sessions. In White Nile, however, there were four communities in which the only facilitator left in November 2014, and the new facilitator(s) started in January 2015, leaving a one to two-month interval during which learning sessions were not supervised.

4.1.5 Staggered approach

A staggered approach was used to support the start in each community by the observers and supervisors. In this way, the technical issues that arose in the first community could be solved before the other communities started. The control groups were tested later.

4.1.6 Iterative process

Due to the iterative development process used, two updates of the game had to be installed during the pilot period. The progress in the communities in North Kordofan, who started first, was faster than anticipated: they had to wait two weeks before the first update could be installed. Furthermore, two mini-games did not function properly which made it impossible to give the right answers. These bugs were fixed within two weeks.

4.1.7 Oral test

Oral mathematics tests were used as all children were assumed to be illiterate (test A & B). These tests were designed on the basis of the Early Grade Mathematics Assessment (EGMA; USAID, 2009), and consisted of 30 items each (maximum score was 60 points), covering the very basics of mathematics. Both tests tested oral counting, number identification, one-to-one correspondence, quantity discrimination, word problems, addition and writing down numbers. In test A the numbers ranged from 1 to 10, in test B the numbers ranged from 1 to 20. The same tests were used as a pre-test and post-test.

4.1.8 Test protocol

As the children live in remote communities, it was assumed they had not been tested before in any formal way. Reports on the testing of children in developing countries mention that children are shy to answer any questions at all (Kanu & van Hengel, 2013). A test protocol was designed, including child-friendly approaches. The observers were trained to use this. During testing, a supervisor was present to ensure that the testing was performed according to protocol.

4.1.9 Data included in further analyses, experimental group

Three children were excluded from the experimental group; one because he was too young (6 years old), two because they had been to school before. Facilitators reported that 57 children dropped out during the pilot. Five more children were excluded on the basis of the logged data; their logged tablet data showed they had only played the game for a short period, and not finished it. The data of the remaining 526 children could be used in further analyses.

4.1.10 *Data included in further analyses, control group*

There were some issues with the data collection in the control group: the pre-test and post-test were not taken at the designated times. In Gedaref, the pre-test of test A was taken two months after the children had started their lessons. Due to this, all data from the Gedaref control group (100 children) was excluded from further analysis. In White Nile and North Kordofan, the pre-test of test A was taken at the right time; the post-tests, however, were taken later than planned. Instead of an interval of 6-8 weeks, the post-test was taken after three months in North Kordofan (45 children) and after six months in White Nile (180 children). Consequently, the results from these two states were analyzed as two different sets. There were no post-test data for 14 children in North Kordofan and for 34 children in White Nile. The data of the remaining 177 children was included in further analyses. Due to logistic issues, logged data were not collected for two communities (57 children).

4.1.11 *Logged data*

The matching of logged data to the test data proved difficult, As not all facilitators had used the unique child numbers for this, but had instead used the child's name. Names in Arabic can be spelled in different ways, which made it impossible to match 23 files of logged data. For 449 of the children in the experimental condition logged data were available and included in further analyses.

4.2 Research instruments

4.2.1 *Geographic information*

The following geographic information was collected by the project team: names of state, locality and community, GPS location of communities, characteristics of community, distance to nearest primary school, and distance to nearest secondary school (see Appendix A).

4.2.2 *Demographic information*

Demographic information from the participating children was collected using a questionnaire for parents. Information on age, gender and family situation (parents, education of parents and number of siblings) of the children was obtained (see Appendix B).

4.2.3 *Educational research*

To assess learning effect, three different types of data were collected. First of all, mathematics tests were developed and carried out in a pre-test post-test control group design. Furthermore, all actions children took in the game were logged on their personal account. Finally, independent consultants carried out EGMA an Early Grade Mathematics Assessment (USAID, 2009) with a stratified sample of the children in the experimental group.

4.2.3.1 *Mathematics tests*

To assess progress of children with respect to mathematics, four different mathematics tests were developed (see Appendix C). Each test consists of 30 items. A good answer straight away receives 2 points; a good answer when asked a second time receives 1 point; a wrong answer receives no points. Children can score a maximum of 60 points on each test.

All tests were taken as a pre-test and a post-test. The tests were developed on the basis of EGMA. The focus of EGMA is on the early years of mathematics learning, with an emphasis on numbers and operations and on geometry through second grade or, in developing countries, perhaps through third grade.

The tests are related to the subjects the children can learn by playing the game. They have different difficulty levels, with test A being the easiest and test D the most difficult. The most important difference with EGMA is that there is no time-limit in these tests. All tests have been approved by the National Council of Mathematics in Khartoum.
Test A: numbers up to 10; number discrimination up to 10; missing number up to 10; word problems, addition up to 10; addition up to 10.
Test B: numbers up to 20; number discrimination up to 20; missing number up to 20; word problems, addition up to 10; addition up to 10.
Test C: number discrimination up to 20; missing number up to 20; word problems, addition up to 20; addition up to 20; word problems, subtraction under 10; subtraction under 10; shapes.
Test D: word problems, addition up to 20; addition up to 20; word problems, subtraction under 20; subtraction under 20; shapes.

4.2.3.2 Logged data

All actions the children took in the game were logged on a tablet in their personal user account. As the internet connection in the villages was not reliable, the data could not be synchronised to a server. At the end of the pilot, these data were downloaded onto a laptop, and made available for analysis (see Figure 4).
The following data were collected per child: number of days played, percentage complete, number of mini-games started, number of mini-games finished without any mistakes. In addition, data on the mini-games and instruction videos were collected: which mini-game was started most often; which instruction video was watched most often?

Figure 4 ‘Milking the tablets’ in El Dabaseen.

4.2.3.3 EGMA

In addition to the project-internal testing of mathematical knowledge, external consultants were contracted to take the Sudanese version of EGMA with a stratified sample of the children (210 children). A separate report of this research was provided by the consultants themselves (Adel Abdullah Idris Siddiq, 2015); a summary of their findings is included in this report as well.

EGMA consists of ten subtests, each measuring essential mathematical concepts that children need to know before they can continue with more difficult mathematical concepts.

1 Number identification
Children are asked to identify randomly-selected printed numbers.
2 Quantity discrimination
Children are presented with pairs of numbers. They have to identify which number is higher than the other.
3 Missing number
Children are presented with a series of 4 numbers. Three number are given, the fourth one is missing. They have to identify which number is missing.
4 Addition 1
Abstract additions up to 20.
5 Addition 2
Abstract additions up to 100.
6 Subtraction 1
Abstract subtractions up to 20.
7 Subtraction 2
Abstract subtractions up to 100.
8 Word problems
Children have to solve mathematical problems that are presented in short stories. The mathematical concepts are: addition, subtraction, multiplication and division.
9 Shapes 1
Children are presented with shapes (square, rectangle, circle, and triangle). They have to identify the right shape.
10 Shapes 2
Children are presented with a number of shapes, in a specific order. Children have to identify the pattern in which the shapes occur.

4.2.4 Psychosocial research

To explore and establish psychosocial impact of learning and ICT on the children and the communities they live in, the children answered an oral psychosocial questionnaire twice (see Appendix D). The development of the psychosocial questionnaire was described in more detail in Stubbé, van der Klauw & Langefeld (2014), this report provides a short summary. In addition, there were focus group meetings for children and for parents to gather more qualitative data about psychosocial effects.

The constructs used in the questionnaire were: Self-esteem, Self-efficacy, Motivation to learn, Social support, Future orientation, and Identity orientation. These construct were chosen as they are known to interact with learning outcomes.

Definitions

(1) Self-esteem: global self-esteem is typically defined as one's overall sense of worthiness as a person (Schmitt & Allik, 2005);
(2) Self-efficacy: one's belief in one's ability to succeed in specific situations (Bandura, 1977);
(3) Learning motivation: intrinsic motivation to learn, learning for its inherent satisfaction rather than for some separable consequence (Ryan & Deci, 2000);
(4) Social support: the support children feel they receive from their parents, family and community (Fleuren, Paulussen, van Dommelen & van Buuren, 2012);
(5) Future orientation: the ability to recognize potential in the form of future possibilities and alternative choices (Adams & Marshall, 1996), and
(6) Identity orientation: people derive their sense of self (identity) largely from the social categories to which they belong, this is unique per person (Adams & Marshall, 1996).

Self-esteem was measured using four pictures of a tree (see Figure 5). These trees relate to the 'tree of life' that has been used for psychosocial research in Sudan before: children are asked to draw a tree that reflects how they feel about themselves. The drawings are then used to assess the level of Self-esteem.
In the current study the four trees represent a four-point Likert scale. Children were asked to point at the tree that showed how they felt about themselves, the bare tree meaning they do not feel too good about themselves and the full tree with flowers meaning they feel very good about themselves.

Figure 5 Pictures of trees, 4-point Likert scale for Self-esteem.

The other constructs in the psychosocial questionnaire were measured with five cups, ranging from an empty cup to a completely filled cup (see Figure 6, below). This represents a five-point Likert scale. Children were asked to respond to statements by pointing at the cup that suited their situation best. The empty cup means the statement does not match their feelings at all, the completely filled cup means the statement matches their feelings very well.

Figure 6 Picture of cups, 5-point Likert scale for psychosocial questionnaire.

4.2.5 Game interaction research

To explore the interaction of children with the game, and assess their feelings about the game, the children answered a questionnaire at the end of the pilot. Children responded to the statements using the same 5-point Likert scale with the five cups described above. The statements were based on the questionnaire used in the Game-based learning Evaluation Model (Oprins, Visschedijk, Bakhuys-Roozeboom, Dankbaar, Trooster & Schuit, 2015). The questionnaire is included in Appendix E.

In addition, children were asked to answer three questions per mini-game:
1. Did the mini-game work?
2. Did you know what you had to do in this mini-game?
3. Did you like this mini-game?

These questions were answered using the 5-point Likert scale with cups, described above.

4.3 Research plan

Table 2, below, shows an overview of how the tests and questionnaires were to be used during the pilot. T stands for the time in weeks. T=0 means before the pilot starts. A staggered approach was used: the communities do not start on the same date. Testing is supposed to be done according to the number of weeks the children have been learning with the game.

Table 2 Research plan Pilot II.

Instruments \ time	T=0	T=7	T=14	T=20	T=26
Demographic information	X				
Geographic information	X				
Mathematics test	A-PRE	A-POST B-PRE	B-POST C-PRE	C-POST D-PRE	D-POST
Psychosocial questionnaire	X		X		X
Psychosocial focus group meeting				X	
Log files tablets					X
Game interaction questionnaire					X
EGMA					X

Due to logistics, the research was not completely carried out according to the research plan. As a result only two of the mathematics tests were used (test A: T=0 & T=8 and test B: T=10 & T=18), and only two measurements of the psychosocial questionnaire were carried out (T=0 and T=26). All the other research instruments were used, on time.

4.4 Local researchers

Local researchers were trained to gather the data using the mathematics tests and the questionnaires. They were trained in child-friendly approaches, in the background of the psychosocial questionnaire and in the way in which the tests and questionnaires should be taken. By using neutral, trained researchers (instead of facilitators) we made sure that there was no bias when administering the tests and questionnaires and that the tests and questionnaire were taken in similar ways with all the children.

4.5 Ethics

- The ethics committee of the Ahfad University for Women has approved these pilot studies.
- Agreements have been signed by all three States and all the participating communities.
- All facilitators have signed the War Child Holland Child Safety Protocol.
- Parents have signed consent forms for their children to take part in the pilot study and to be photographed.
- All data are collected individually and related to a child-specific number (anonymous), see Figure 7, below. This is done for privacy reasons, as well as for pragmatic reasons (Arabic names are spelled in different ways in English).
- The communities in the control group in this pilot, will participate in a later phase of the project.

Figure 7 Recording of child-specific numbers.

5 Results

Figure 8 Rural community in White Nile.

A total of 591 children participated in the baseline study of Pilot II. During the pilot, some of these children missed learning sessions or dropped out altogether.
In defining drop out, we specifically aim to include those children who have participated during the whole pilot period, but may have done so for only a limited number of days per week (2-4 days a week instead of 5). Therefore, we started from the facilitators' observations and remarks. Facilitators reported 57 children who had left the program before it finished. For most children, the reason for drop out was recorded. In addition to the facilitators' remarks, five more children were identified as drop outs: the logged data collected from the tablets showed that they had stopped playing the game before 12 March, and had finished less than 90% of the game. Two other children, who had also stopped playing the game before 12 March were not considered drop outs, because they had finished more than 90% of the game. That means 62 children (10%) dropped out in Pilot II.

In addition, four more children were excluded from further analyses. For one child the logged data and facilitator remarks were inconclusive; it was not possible to determine whether this child had dropped out or not. The other three children took the baseline tests after they had started playing the game. Therefore, their results could not be included in the analyses. Finally, five additional children were excluded from the analyses because they had missed more than two mathematics tests.
For four of these children this was confirmed by facilitator remarks. The fifth child had a hearing problem and could not manage the programme and the (oral) tests.
The data of 517 children was included in further analyses. An overview of the data that were excluded from further analyses is shown in Figure 9.

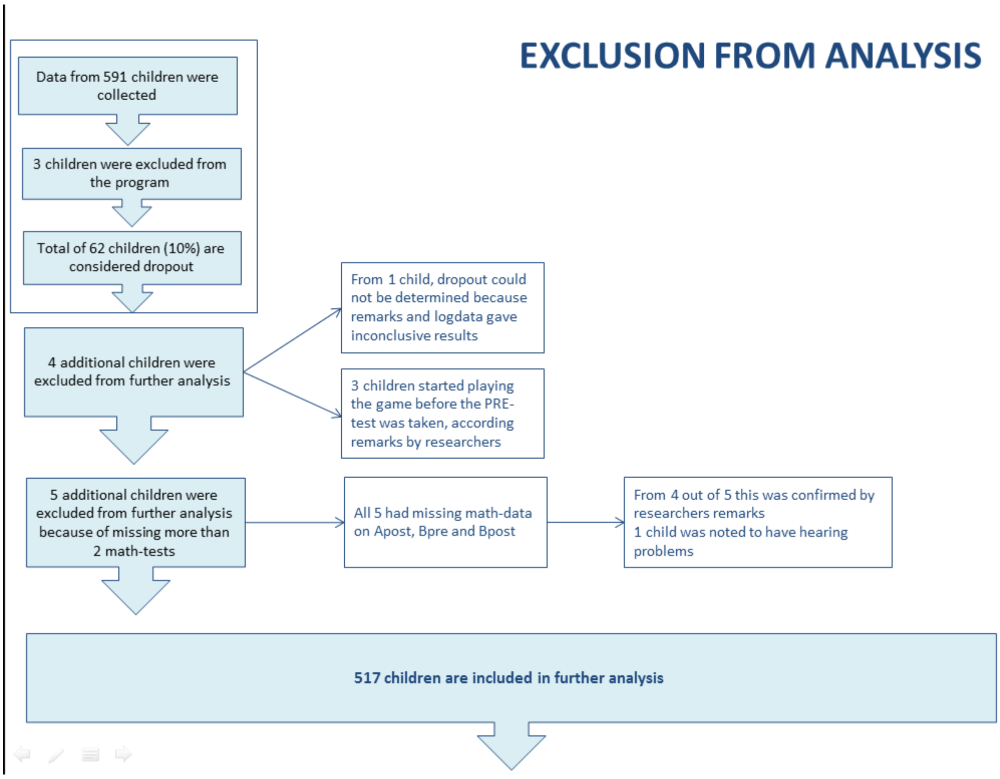


Figure 9 Overview of results excluded from further analysis.

Facilitators report several reasons for drop out: according to facilitator remarks 95% of the children dropped out because they moved to another community (e.g. for harvesting crops or finding water) or went to work in the field (see Figure 10). Based on facilitator remarks, 5% of the children (3 children) dropped out because they, or their parents, 'refused the programme'.

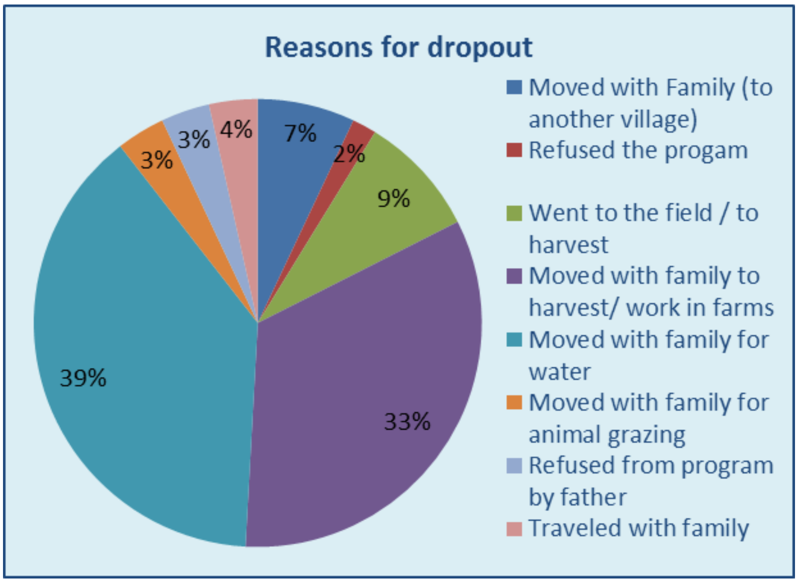


Figure 10 Reasons for drop out.

The characteristics of the children who dropped out do not show any significant differences with the total population. This means that e.g. gender or age is not related to the chance of dropping out. There were significant differences between the states, though. This was caused by the fact that in two communities in White Nile almost all children dropped out. According to facilitator remarks, in these communities many families moved to find water or harvest crops. Also, children who have less siblings drop out more often than children with more siblings (see Figure 11, below).

COMPARING DROPOUTS WITH INCLUDED CHILDREN

		Children considered as Dropouts	Children Included in further analysis
	N:	62	517
	%:	11%	89%
Boy/Girl	Boy	53,2%	48,2%
	Girl	46,8%	51,8%
Age	7	44,3%	47,8%
	8	34,4%	28,2%
	9	21,3%	24,0%
Age	- Mean	7,77	7,76
State	White nile	74,2% ▲	28,6% ▼
	North Kordofan	12,9% ▼	35,2% ▲
	Gedaref	12,9% ▼	36,2% ▲
Siblings	- Mean	4,62 ▼	5,36 ▲
Row	- Mean	1,93	1,91
Mother	finished primary education	11,5%	16,1%
	not finished primary education	88,5%	81,8%
	finished secondary education	0%	2,1%
Father	finished primary education	14,8%	18,0%
	not finished primary education	83,6%	79,5%
	finished secondary education	1,6%	2,5%
	Note. Percentages are column percentages, and are tested with the Pearson χ^2 test (horizontal comparisons). Means are tested with the t-test (horizontal comparisons). The contrast is each subgroup vs all other cases (weighted deviation contrast). ▲: p<0,05 (and ▼): Significant high (low) percentage and/or mean (2-tailed tests).		

Children that are considered dropouts appear to be more often from White Nile and less often from North Kordofan and Gedaref than children that are included in further analysis.
Children that are considered dropouts appear to have fewer siblings than children that are included in further analysis.

Figure 11 Comparing drop outs with included children.

Apart from drop out, there are missing data for some tests for some of the children. The exact characteristics (gender, age and state) for the children with missing data are given in Appendix F.

5.1 Geographic information

The geographic information was described in detail in the baseline study of this pilot (Stubbé et al. 2015). The geographic results are based on all 591 children who were originally enrolled in the study; including the children who dropped out later or were excluded from further analyses.
In North Kordofan seven communities were involved, with a total of 197 children. In White Nile seven communities participated as well, with a total of 195 children. In Gedaref five communities participated, with a total of 197 children. The reason for the smaller number of communities involved was that the community Tayba in Gedaref was a large community with more than 100 children. Figure 12, below, shows the number and names of the communities per state, and the number of participating children per community.

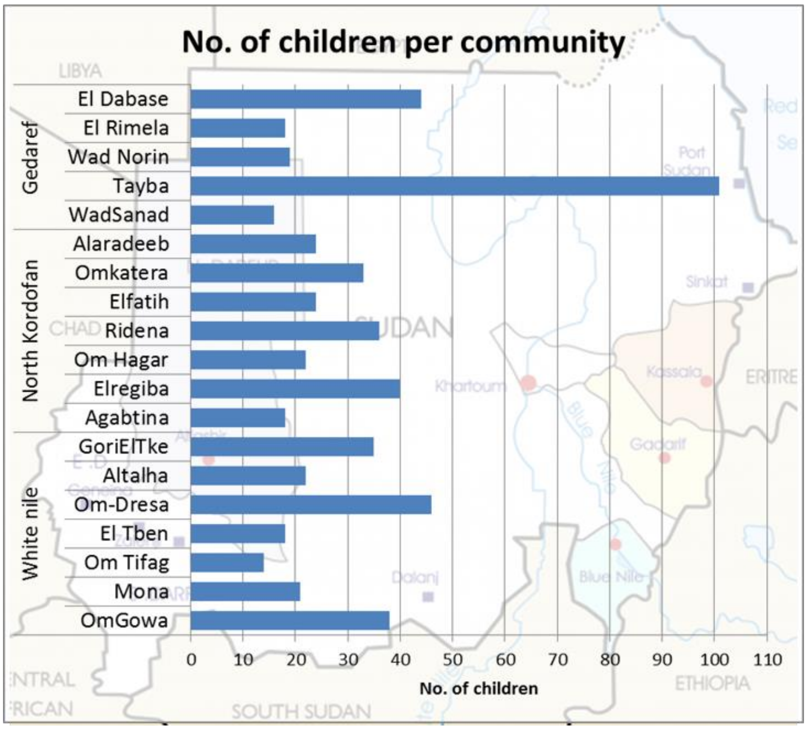


Figure 12 Number of children per community.

The distances to the nearest primary schools varied between three and 14 kilometres, with an average of 7.5 kilometres. The average distance in Gedaref was 6.3 kilometres, and in North Kordofan 6.7 kilometres. The average distance to the nearest primary school was largest in White Nile: 9.3 kilometres.

5.2 Demographic information

The demographic information was described in detail in the baseline study of this pilot (Stubbé, van der Klauw & Langefeld, 2014). The demographic results are based on all 591 children who participated in the study. 51% of the participating children were girls, 49% of them were boys. The participating children were between 7-9 years old. Most of them (47%) were seven years old. 29% was eight years old and 24% was nine years old. The average age of the participating children is 7.8 years. Children in North Kordofan were slightly younger (average 7.4 years), compared to the children in Gedaref (average 8.0 years) and White Nile (average 7.9 years).

Most children reported they had both their parents. Children from North Kordofan reported more often that they only had a mother (24%) or no parents at all (1%).

On average, children were reported to have 5.3 siblings. In North Kordofan this was less, with 4.8 siblings on average, whereas children from Gedaref reported 5.8 siblings per child on average. Most of the participating children were second in the row of children in their family. About 15% of them were the oldest, about 5% were the third child in their families.

Only 20% of the parents has finished primary education, of which only a few have finished secondary education as well. There was no significant difference between fathers and mothers with respect to the level of education.

5.3 Educational research

The educational research is based on three different types of research instruments: mathematics tests carried out by the research team, logged data and EGMA, carried out by independent consultants. In this paragraph the results of these three research instruments are described.

5.3.1 Mathematics tests

Two different mathematics test were used in Pilot II: test A and test B. Each test had a 6-8 week interval between pre and post-test.
Table 3, below, shows the number of children that had missing data either for test A (PRE or POST) or test B (PRE or POST), or for both tests. Data for a test is considered missing when a child did not take one or both of the tests (PRE and/or POST). The reason for this is that both PRE-test and POST-test are needed to assess progress in the test.

Table 3 Missing data mathematics tests.

A-PRE	A-POST	B-PRE	B-POST
2	13	16	47

5.3.1.1 Results test A

Figure 13 Taking test A-PRE in Om Hagar.

To assess if children have increased their scores on a mathematics test an Anova repeated measures (SPSS GLM) test within subjects factor: Math-A-PRE en Math-A-POST was used. The average score of children in White Nile, North Kordofan and Gedaref on the pre-test of test A is 20 (max. 60). The average score on the post-test of test A is 41. The analysis shows that children have increased their scores on test A significantly (F(1,499)=1170.929; p< .001); r=.85.).
There is a significant interaction between math score and state F(2,499)=9.055; p < .001; r=.21) and a significant main effect of state (F(2,99)=21.710; p < .001; r=.29); White Nile has a higher score than North Kordofan and Gedaref on the pre-test (24 points resp. 16 resp. 19) as well as on the post-test (47 points resp. 40, resp. 37) of test A (see Figure 14). Posthoc tests (Bonferroni) show that White Nile has scored significantly better (p < .001) compared to the other two states. Figure 14 also shows that all three states perform better on the Post-test than on the Pre-test, but also shows that North Kordofan has a greater increase of scores from pre-test to post-test.

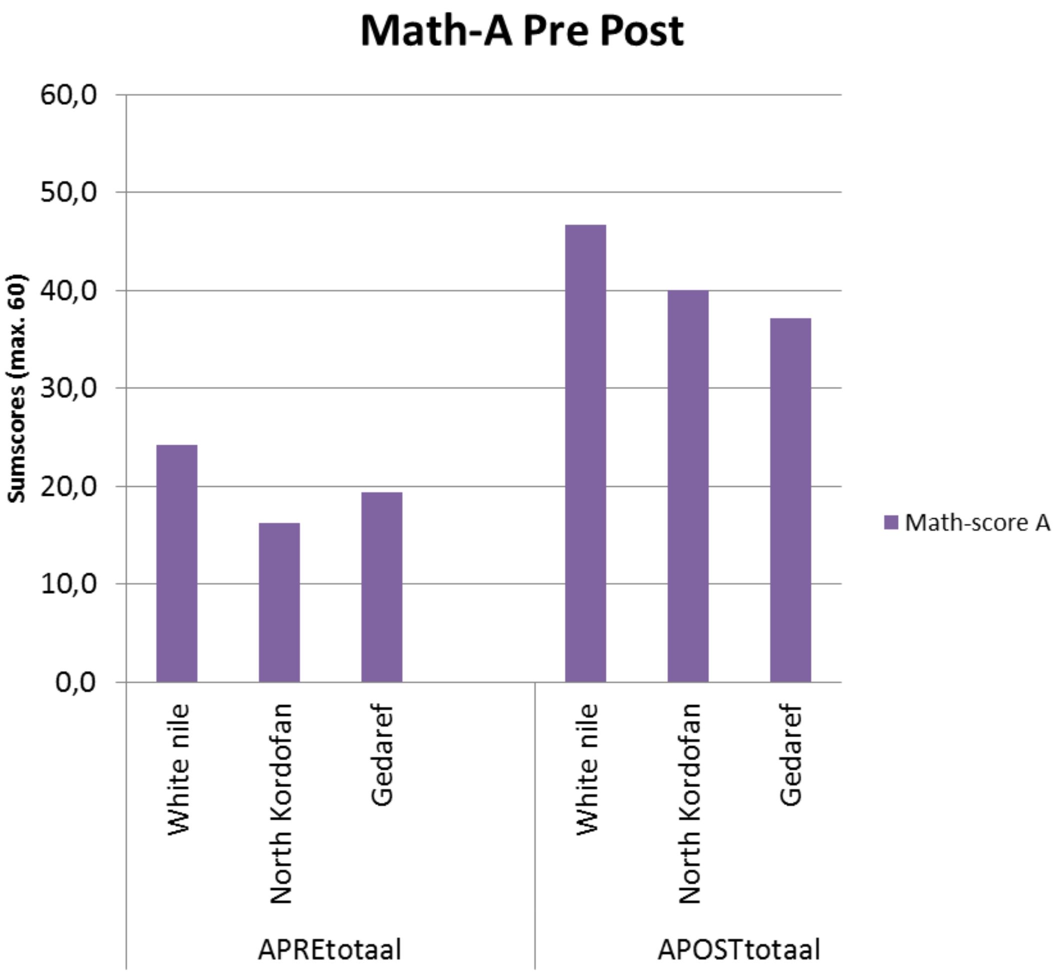


Figure 14 Mathematics scores on test A, per state.

There are no significant differences between boys and girls on either pre-test or post-test of test A (see Figure 15).

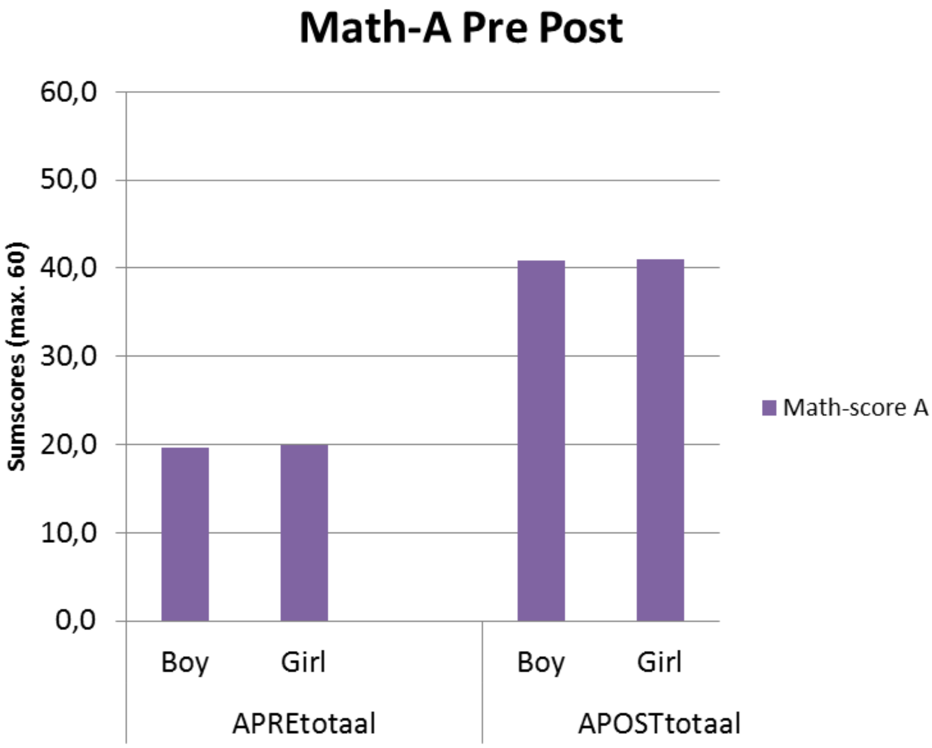


Figure 15 Maths scores on test A, per gender.

There are significant differences between the age-groups, with the younger children scoring lower than the older children ($F(2,499)=14.758$; $p < .001$); $r=.25$). Post hoc analysis shows that the 9-year old children differ from the younger ones. There is no significant difference between the 7- and 8 years ($p = .075$) (see Figure 16).
In addition, there was a significant interaction between math score and age ($F(2,499)=5.258$; $p < .010$; $r=.16$): the younger children show a larger learning effect than the 9-year old children (Figure 16).

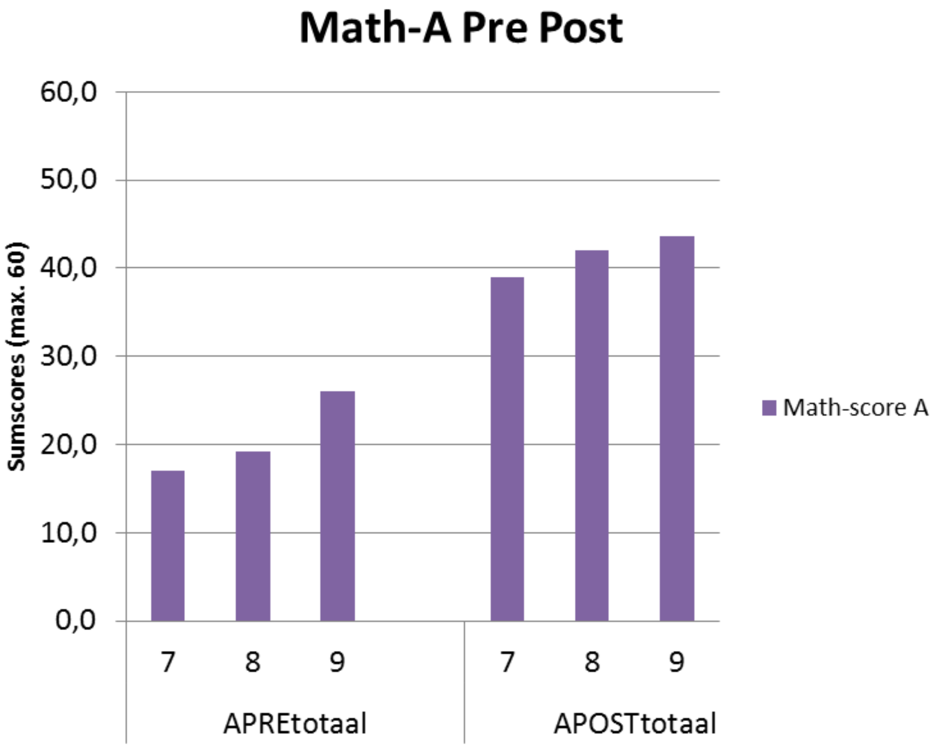


Figure 16 Mathematics scores on test A, per age group.

5.3.1.2 *Results test B*

To assess if children have increased their scores on a mathematics test (between week 8 and 14) an Anova repeated measures (SPSS GLM) test within subjects factor: Math-B-PRE en Math-B-POST was used. The average score of children in White Nile, North Kordofan and Gedaref on the pre-test of test B is 32 (max. 60). The average score on the post-test of test B is 41. The average increase of 9 points is significant ($F(1,456)=160.067$; $p < .001$; $r=.51$). There are also significant differences between the states (see Figure 17) ($F(2,456)=11.967$; $p <.001$; $r=.22$); Post-hoc shows that Gedaref differs significant (Bonferroni) ($p < .010$) from the two other two states, with a smaller learning effect. In addition, there is a significant interaction effect between state and learning effect ($F(2,456)=10.022$; $p<.001$; $r=.20$); the learning effect was highest in North Kordofan.

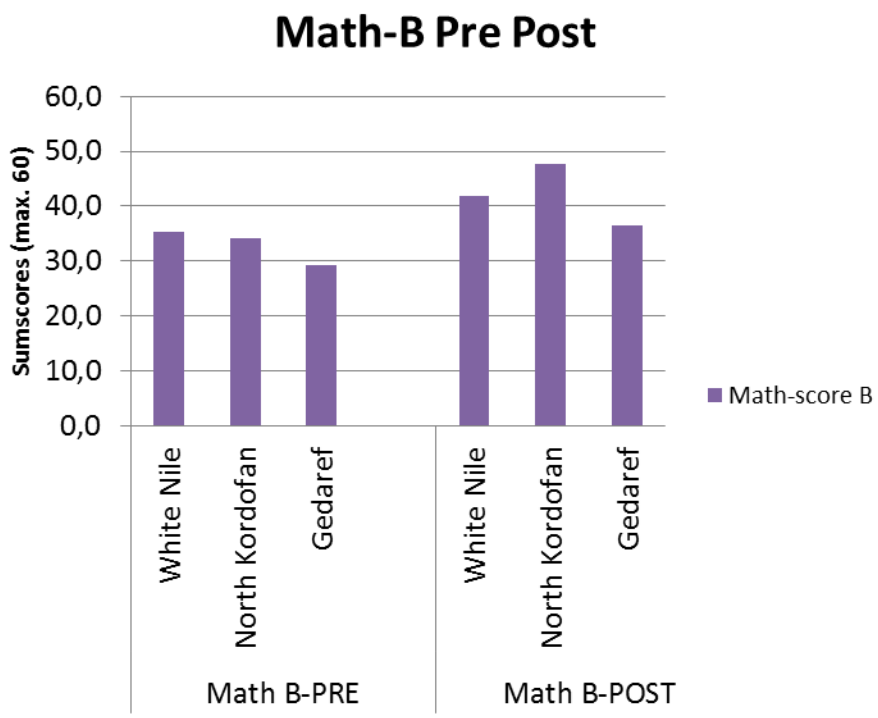


Figure 17 Mathematics scores on test B, per state.

There are no significant differences between boys and girls (see Figure 18).

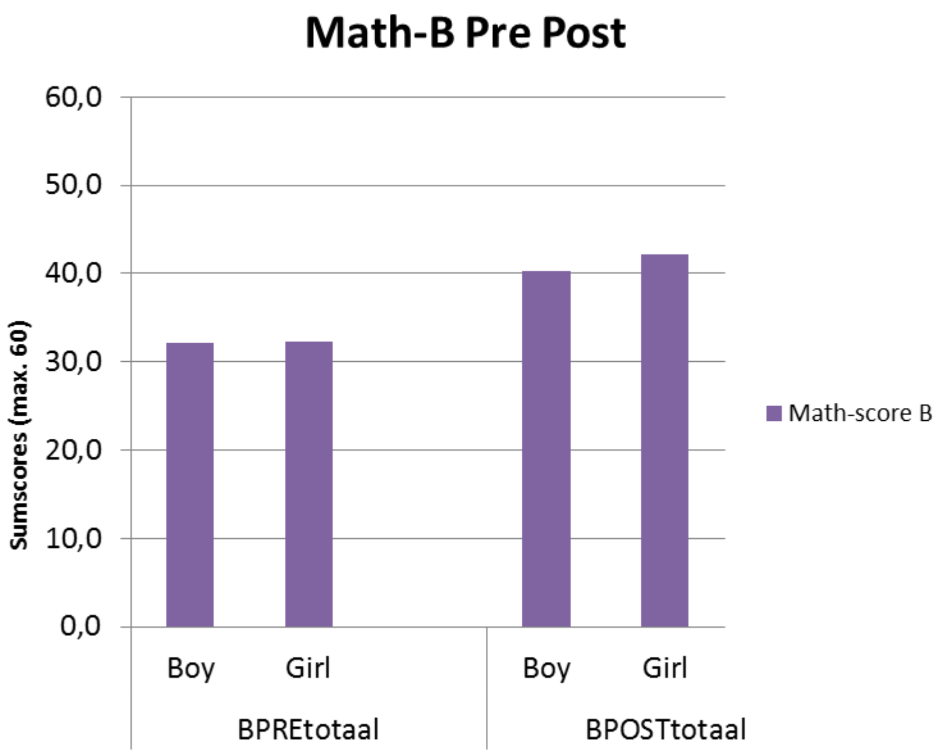


Figure 18 Mathematics scores on test B, per gender.

There are no significant differences between the age-groups (see Figure 19).

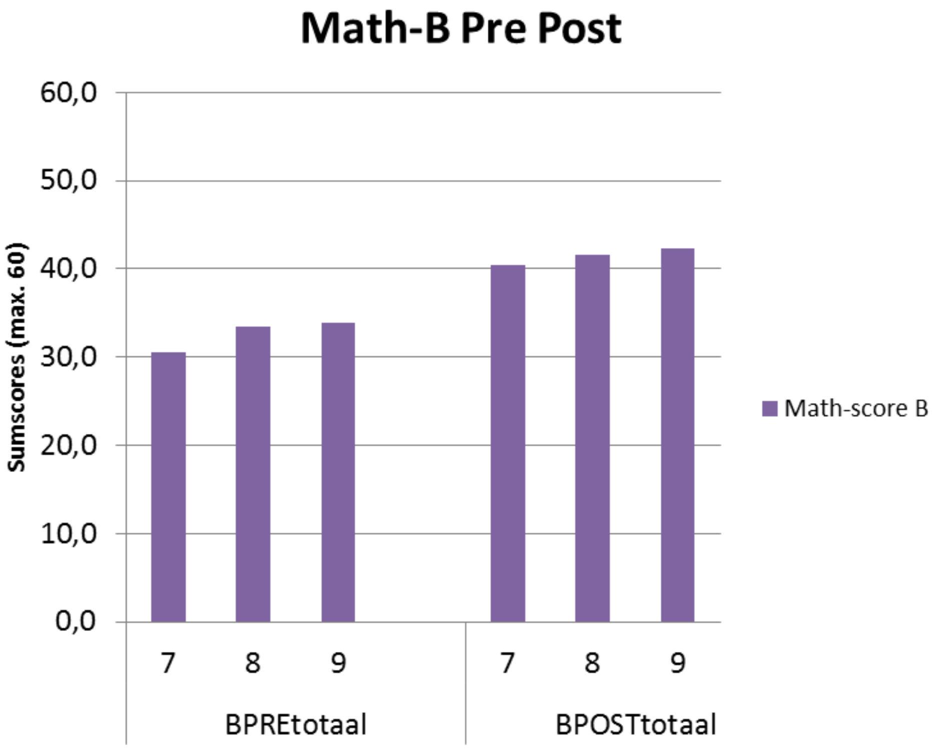


Figure 19 Mathematics scores on test B, per age-group.

5.3.1.3 *Results Control group*

Due to the differences in the order of the curriculum, only test A could be taken with the control group; they had not been taught the mathematical concepts of test B. As described before, there were some issues with the data collection in the control group: the pre-test and post-test were not taken at the designated times.
In Gedaref, the pre-test was taken two months after the children had started their lessons. Due to this, all data from the Gedaref control group (100 children) was excluded from further analysis. In White Nile and North Kordofan, the pre-test was taken at the right time; the post-tests, however, were taken later than planned. Instead of an interval of 6-8 weeks, the post-test was taken after three months in North Kordofan (45 children) and after six months in White Nile (180 children). Consequently, the results from these two states were analyzed as two separate sets. There were no post-test data for 14 children in North Kordofan and for 34 children in White Nile. The data of the remaining 177 children was included in further analyses.

Due to the data collection issues with the control group regarding the timing of testing, differences between the experimental group and the control groups must be interpreted with caution. Comparisons are made per state, not with the total average scores.

A comparison of the North Kordofan (NK) experimental group with the NK control group shows no significant differences, in pre-test, post-test or in increase of scores between pre- and post-test (see Figure 20). This is a positive finding, as the NK control group received twice as much instruction per day, compared to the experimental group (2 times 45 minutes informal education, vs. playing the game 45 minutes a day). Moreover, the NK control group had a three month interval between pre and post-test, whereas the experimental group had a 6-8 week interval between tests. Roughly, the NK control group has had three times as much opportunity to learn as the NK experimental group. The NK control group is very small (N=31), which makes it less suitable to compare to the NK experimental group (N=182).

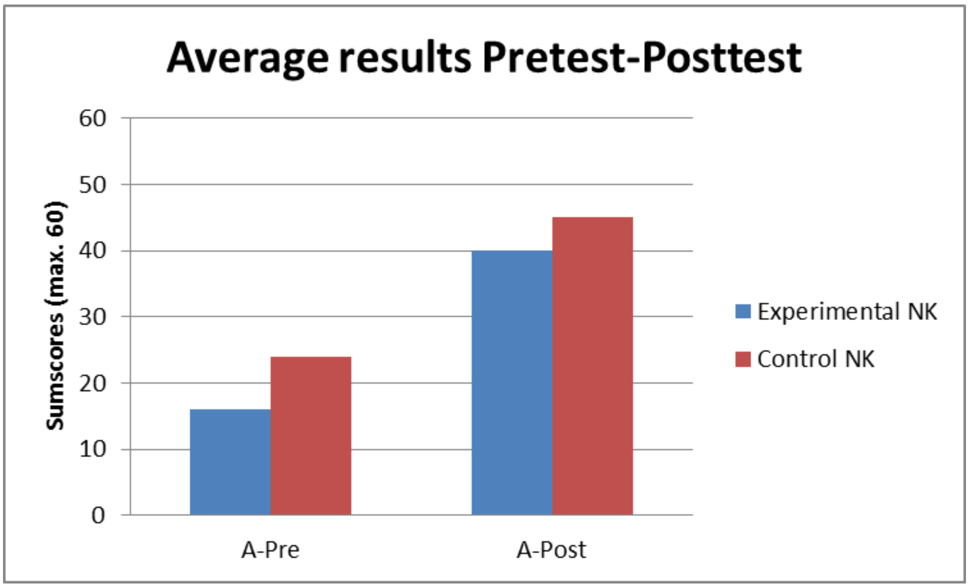


Figure 20 Average results test A, Pre-test-Post-test North Kordofan, control and experimental group.

A comparison of the White Nile (WN) experimental group (N= 148) to the WN control group (N=146) shows that there is a significant difference in increase of scores (31 vs 36 points) between the two groups ($F(1,288)=17.034$; $p < .001$; $r=.24$). The WN control group on average significantly increased its score more than the WN experimental group (see Figure 21).

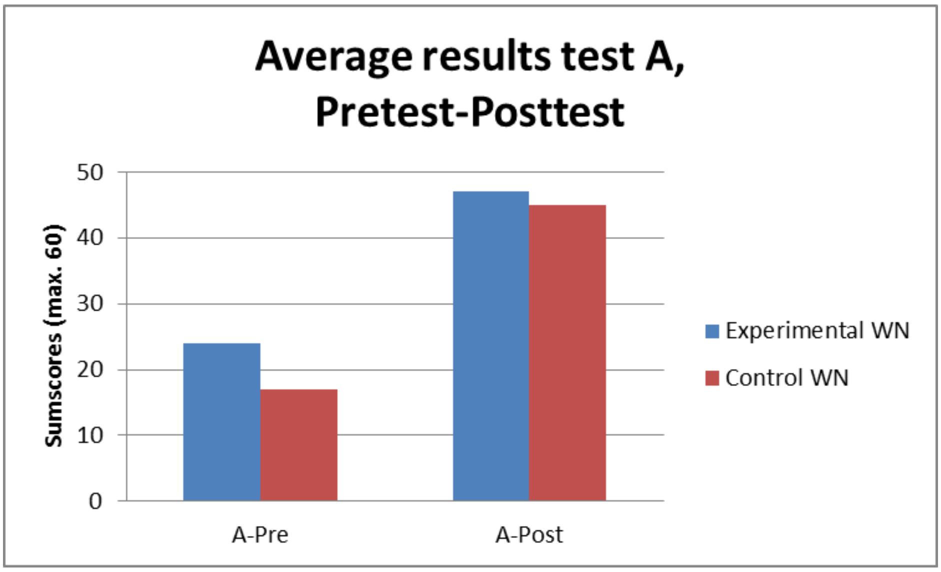


Figure 21 Average results test A, pre-test-Post-test, White Nile, control and experimental group.

It is important to note that the WN control group had (much) more opportunity to learn than the experimental group: they had two mathematics lessons of 45 minutes per day versus 45 minutes in the WN experimental group, and a much longer interval between pre-test and post-test (6 months vs. 6-8 weeks). This means the WN control group had roughly six times more learning time than the children in the WN experimental group. The difference between the experimental and the control group is significant, but rather small. The WN control group had 75% correct on the post-test, which means that there was more room for improvement.

5.3.2 *Logged data*

Logged data of 532 tablets was collected. Due to logistic issues, logged data were not collected for two communities (57 children). The matching of logged data to the test data proved difficult, because not all facilitators had used the unique child numbers to create the accounts, but had instead used the children's names.
386 accounts were based on the child's name written in English, the other 146 were based on the child's name written in Arabic. The Arabic names were translated into English by a native Arabic speaker, and all the English names were then matched to the original names collected during the baseline. The matching was performed by three researchers, individually. Only the names that were matched by all three researchers were included in the final list. As names in Arabic can be spelled in different ways, it was impossible to match 23 files of logged data to names of children from the baseline study. At the same time, there were no matched logged data for 83 children. 27 of these were drop outs, that leaves 56 included children without matched logged files (see Table 4, below). For 449 of the children in the experimental condition logged data were available and included in further analyses.

Table 4 Overview missing data logged data.

Baseline study	**Logged data**	**Matched logged data**	**Children without match logged data**	**Logged data not matched**
591	532 (91%)	508 (86%%), including drop outs	83 (14%), including drop outs	23

Most of the children played the game for a period of 5 to 7 months (average 135 days). Girls are found to participate for a longer period than boys (average of 141 versus 129 days; this difference is significant ($F(1,378) = 7,726$; $p < 0,05$).
No differences are found for age. Children from North Kordofan played a significantly shorter period (122 days), compared to White Nile (138 days) and Gedaref (145 days ($F(2,377) = 11,114$; $p < .001$). Correcting for both gender and age, children were found to only differ in the number of days between first and last play based on state ($F (2,363) = 6.123$; $p < .05$).

During the pilot period, most children are found to participate two to three times a week on average (total number of days played divided by the number of weeks between first and last time played per child). Frequency of playing is an average that can be the result of playing five times a week for a number of weeks and then skipping one or more weeks, or playing a limited number of days per week. Figure 22 represents the frequency of playing (average plays per week), in categories.

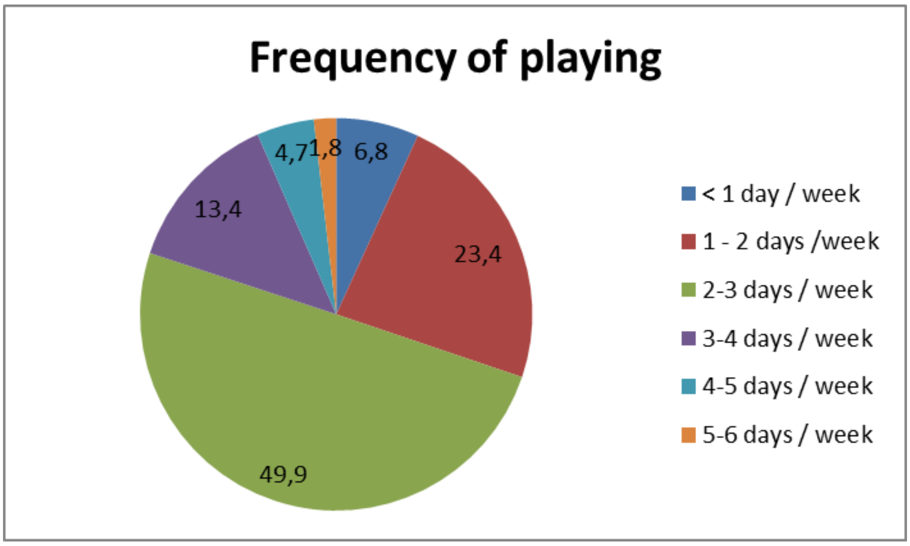


Figure 22 Frequency of playing.

No differences are found for gender. Age is significantly related to frequency of playing: 7-year olds played more often (2,7 days a week) than 8-year olds (2,3 days a week), who in turn played more often than 9-year olds (2,0 days a week; $F(2,377) = 17,909$; $p < .001$). There was also a difference between states: children from North Kordofan played more often (3,2 days a week) than children from White Nile (2,2 days a week), who in turn played more often than children from Gedaref (1,7 days a week; $F(2,377) = 143,954$; $p < .001$). Correcting for both gender and age, children were found to only differ in the frequency of playing based on state ($F(2,363) = 80.093$; $p < .001$) (see Figure 23).

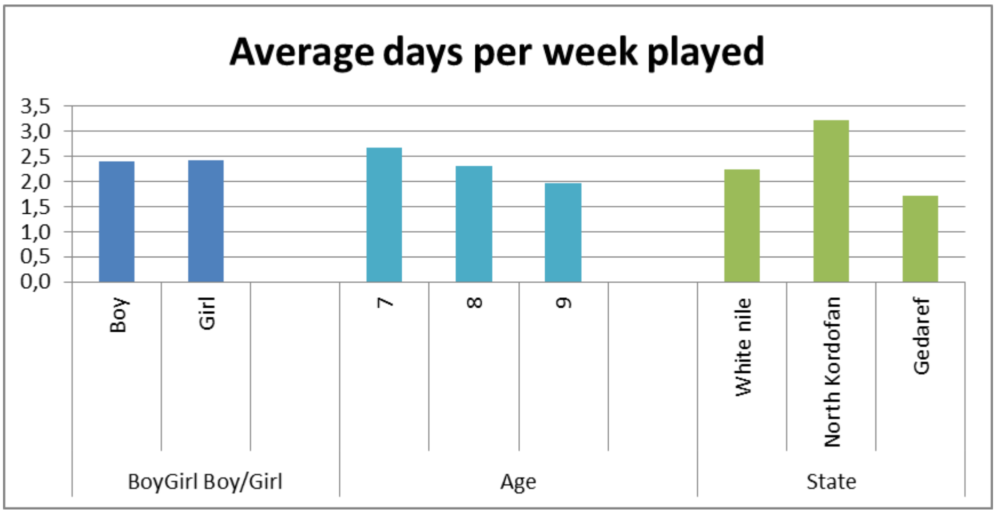


Figure 23 Average days per week played, per gender, age, state.

Children have played an average number of 2592 games (max. 8347), which is an average of 19.2 games per learning session (45 minutes). Children do not differ in the number of games played based on gender. However, children of age 7 tend to play more games than children of age 8 and 9 and children from Gedaref tend to play less games than children from the other states. This is probably related to the average number of days played.

'Percentage complete' was calculated using the six-month curriculum developed for this pilot. As the game has never been played before, we had to make an educate guess regarding how much of the curriculum needed to be included in the game to support six months of learning. It is possible we included too much or too little of the curriculum, which influences the percentage complete.

As shown in Figure 23, above, most children played an average of 2-3 days a week. Nevertheless, 77% of the children completed 50% or more of the game (see Figure 24). 57% of the children completed 90% or more of the game.

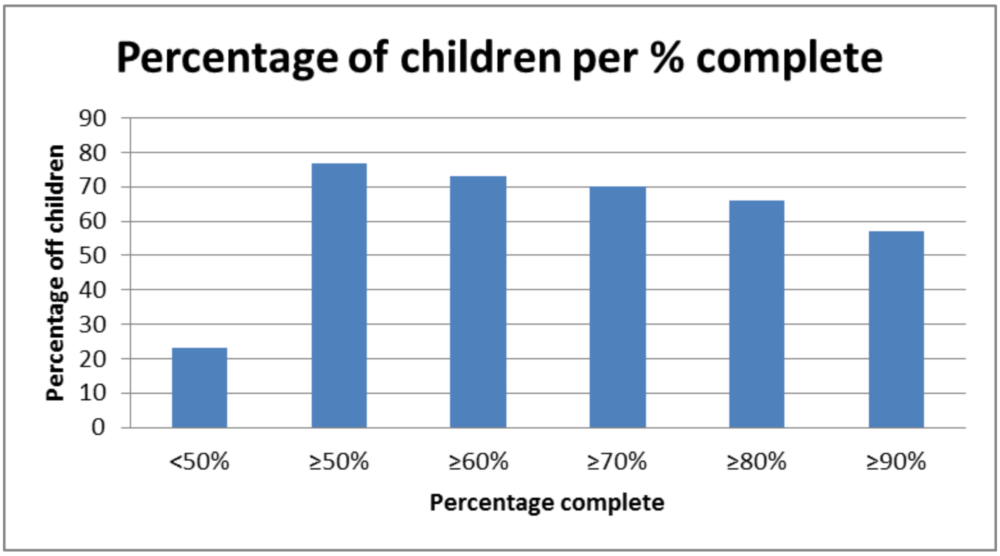


Figure 24 Average percentage of children per percentage complete.

Children who played for five months or longer, had a significant higher percentage complete than children who played four months or less. At the same time, children who played two days per week on average, or more, had a significant higher percentage complete than children who played less than an average of two days per week (see Figure 25). There is an interaction effect between Period and Frequency: when children played less than four months, the effect of frequency of playing on percentage complete is larger than for children who played five months or more. This means that the effect of frequency of playing for children decreases when children play for five months or longer. There are no significant differences between children who played for a short period of time with a high frequency or children who played for a long period of time with a low frequency.

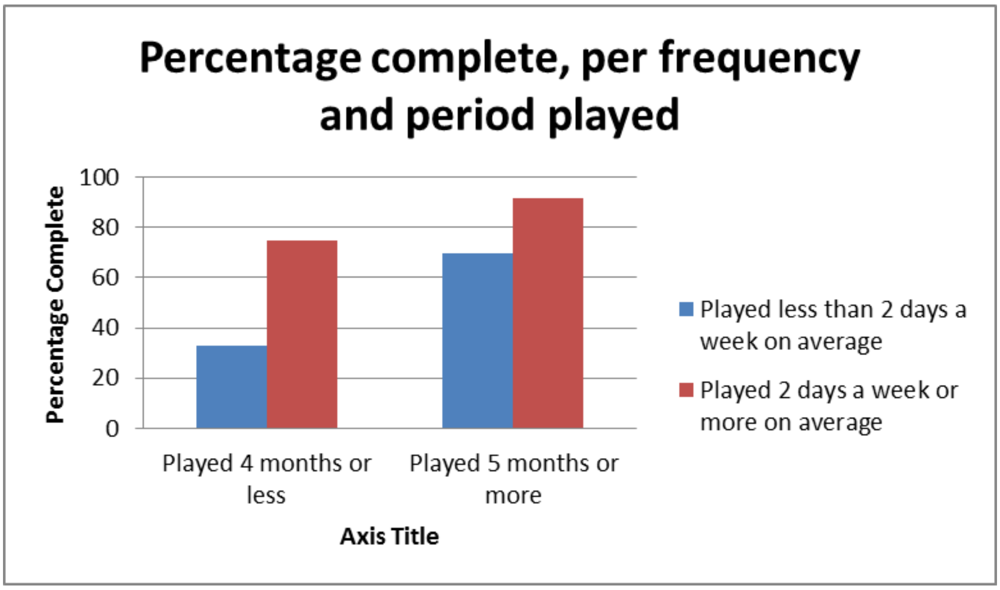


Figure 25 Percentage complete, per frequency of playing and per period played.

5.3.3 *Results EGMA*

Figure 26 Taking EGMA in Mona.

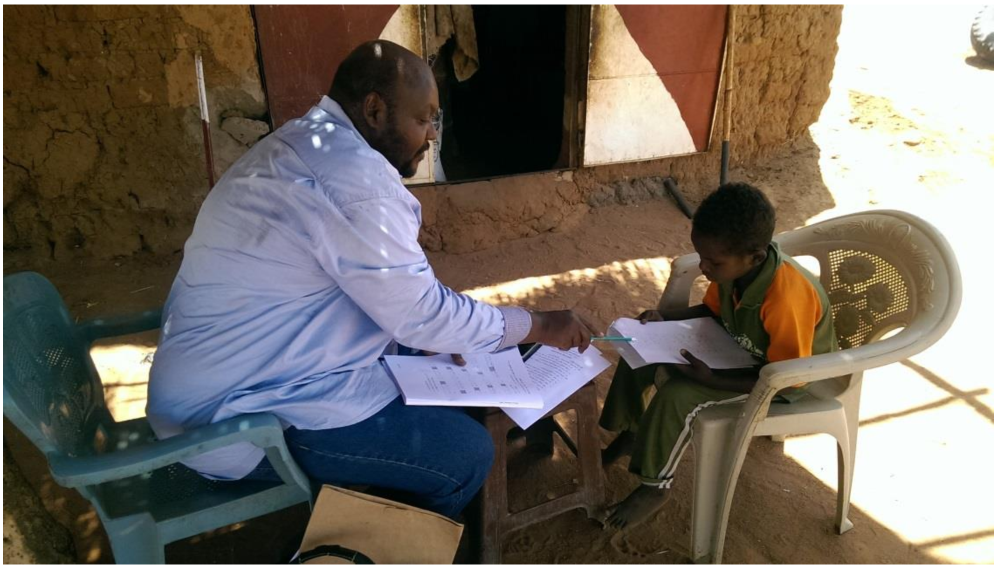


Figure 27 Taking EGMA in Al Draisa.

The results of EGMA, taken by a stratified sample of 210 of the children in the experimental condition, and carried out by independent consultants, can be compared to earlier studies with Arabic speaking children in Khartoum and Jordan, because it is an internationally validated test. Figure 28, below, shows that the children in the experimental condition (ELS) had the highest percentage correct in three sub-tests of EGMA (Shapes I, Shapes II and Word problems) after only six months of learning, compared to children who had attended school for 2.5 years in Khartoum and Jordan. In a fourth sub-test (Missing number) the children in the experimental condition had a higher score than the children from Khartoum.
The children in the experimental condition only had a slightly lower score than children from Khartoum on: Number discrimination, Addition level 1, and subtraction level 1.

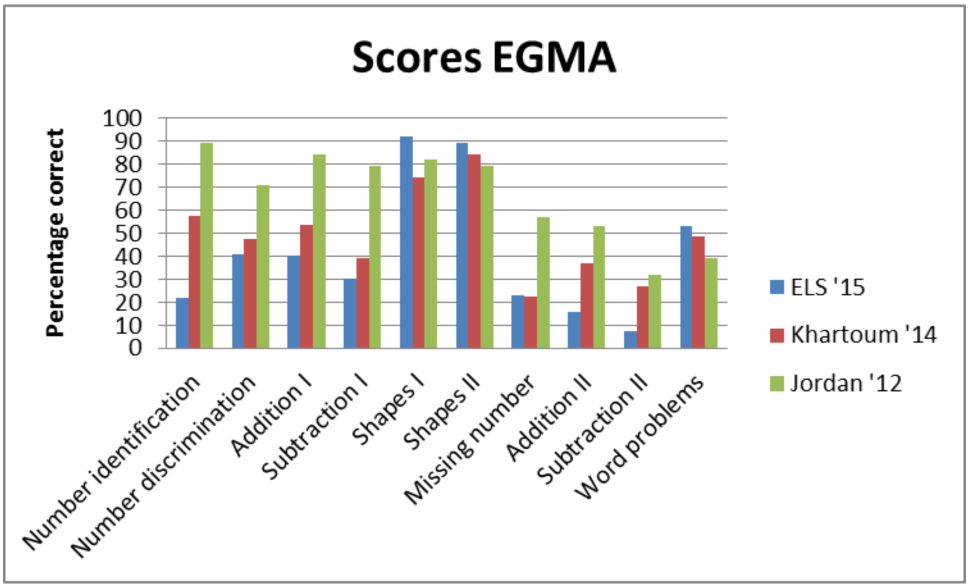


Figure 28 Scores EGMA compared; E-Learning children (ELS), Khartoum and Jordan.

On three sub-tests of EGMA the children in the experimental conditions had quite low scores compared to the children from Khartoum and Jordan. The reason for this is that these subjects had not been taught yet: in EGMA Number identification covers the numbers up to 1000. The children playing the game had only been taught the numbers up to 20. Addition II and Subtraction II cover additions and subtractions over 20. The children playing the game had only been taught addition and subtraction up to 20.

Overall, the feeling of the consultants was that the children in the experimental condition did very well. There were no significant differences for gender. For age, there was a significant difference for the measurement Problem solving and Addition level 2; the 7-year olds had a significant lower score than the 8 and 9-year olds.

5.3.4 *Correlations between educational research instruments*

The educational research was based on three different types of research instruments: (1) mathematics tests developed and used within the project team, (2) logged data from the tablets, and (3) EGMA, an internationally validated mathematics test carried out by independent consultants.

5.3.4.1 *Correlations between Mathematics tests and EGMA*

There are significant positive correlations between the scores on the mathematics tests A and B (pre and post) and the measurements of EGMA. These correlations are not always strong. The strongest correlations are for Addition level 1, Subtraction level 1 and Problem solving, most correlating more than .50 (see Table 5).

Table 5 Correlations between test A and B and measurements of EGMA.

	A-PRE	A-POST	B-PRE	B-POST
Addition level 1	.413**	.504**	.542**	.599**
Subtraction level 1	.502**	.490**	.558**	.538**
Word problems	.486**	.546**	.514**	.611**

**. Correlation is significant at the 0.01 level (2-tailed).

For Addition level 1 and Word problems, the correlations were strongest for B-POST. For Subtraction level 1 the correlation was strongest for B-PRE, but also strong for B-POST.

5.3.4.2 *Correlations between Mathematics tests and logged data*

The percentage of the game children have completed can be seen as a level of competency (mastery). Since the children can only progress to the next level once they have mastered the previous one, the level they have reached is an indication of mastery.

The higher the percentage of the game the child has completed, the higher the scores on the mathematics tests ($F(25,430) = 4952,837$; $p < .000$). Test results are better for children who participated for a longer period of time, even if their frequency of playing was lower. Children who played the game more often (total number of days played) do not necessarily have better results on the mathematics tests.
This means that playing the game in itself does not influence mathematics results. Children are allowed to make one mistake per mini-game and still continue to the next level. Children who finish more mini-games without any mistakes have better results on the mathematics tests ($F(6,920)=1347,830$; $p < .009$). This is not caused

by a higher percentage complete, based on the fact that giving only correct answers allows children to move faster through the game, and thus have a higher percentage complete. There is no significant correlation between the percentage complete of the game and the number of games finished without any mistakes.

5.4 Psychosocial research

5.4.1 Psychosocial questionnaire

The psychosocial questionnaire was filled out in week 0 and week 26. Two children missed the baseline measurement and 26 children did not fill out the questionnaire at the end of the pilot.

Figure 29 Drawing while waiting to take the Psychosocial questionnaire in Agabtina.

For the questionnaire, optimal scaling techniques were used, to control for assumptions about interval distributions of the children's answers to each question (Gifi 1990; Spss 2001; Meulman & Heiser 1999). The unidimensionality of the separate scales was evaluated by Principal Component Analysis with Varimax Rotation and by calculating Cronbach's alpha's. Finally, scale scores were obtained by adding optimal scaling quantification scores within scales, and transforming scores linearly to a 0-100 scale. Higher scores indicating more emphasis on the scale's theme.

The overall reliability of the psychosocial questionnaire was acceptable (Crohnbach's alpha = 0.71). Table 6, below, shows that apart from the measurement of self-esteem (4-point Likert scale with pictures of trees), five scales could be identified by factor analysis. The first scale, Self-efficacy, was reliable (Crohnbach's alpha=0.74). The second scale, Motivation and Future Orientation, had a Crohnbach's alpha of 0.63. In the context of this research we accept this reliability (George & Mallery, 2003). The other three scales were not reliable, and were therefore not used in further analyses.

Table 6 Scales Psychosocial research T0 and T26; Cornbach's alpha, number of items and example items.

	Crohnbach's alpha				
	T0 & T26	T0	T26	No of items	Example
Self-efficacy	.74	.71	.77	4	I can handle whatever comes my way
Motivation and Future Orientation	.63	.63	.64	4	I like learning new things
Belong to group	.52	.56	.46	4	I participate in activities in my community
Social Support	.46	.47	.45	3	I feel supported to learn by others in my community
Identity orientation	.56	.63	.47	3	My relationships with people I feel close to is important to me
Total Psychosocial	.71	.67	.74	16	[no example: all items from the psychosocial research together]

Self-esteem is the only scale that showed significant differences between the two measurements (T=0 and T=26); it has increased from 1.9 to 2.5 on a 4-point Likert scale (see Figure 30).

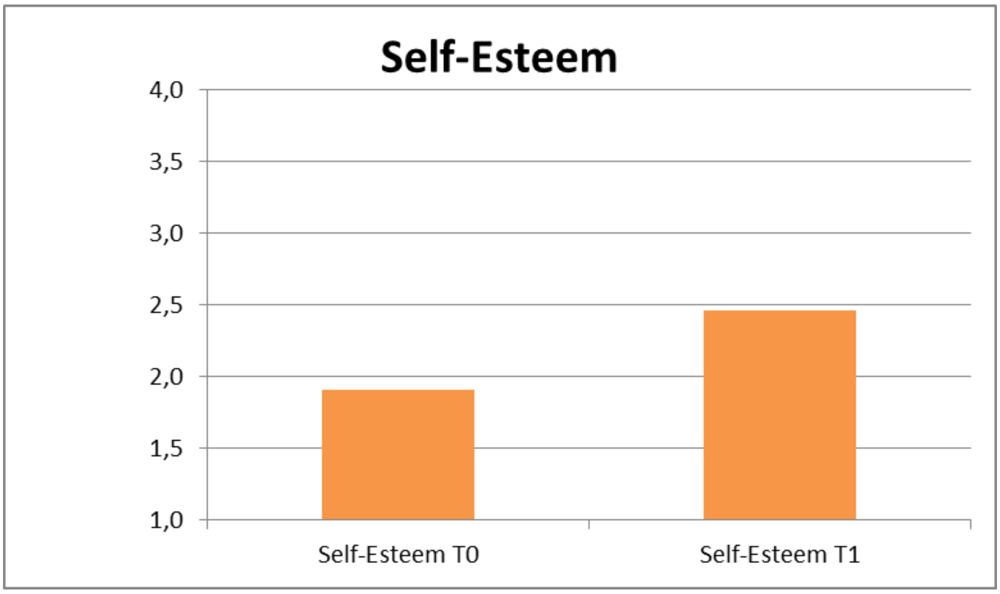


Figure 30 Self-Esteem at T=0 and T=26.

5.4.2 *Focus groups and interviews*

The focus group meetings, with children and with parents, report positive findings. Children mentioned they liked playing the game and are proud of what they have learned. Children indicate they would like to learn more: they would like the project to continue, but also they would like to learn more subjects such as Arabic and English. Children mention they have made friends in the learning sessions.
They now play together or visit each other's huts, which they did not do before.

Parents indicated they feel it is important for their children to learn. For that reason, they mention they would like the project to continue. Parents also mention that they feel their children are safe and looked after during the learning sessions.

In interviews, facilitators mention that children have become more disciplined because of the learning sessions. Facilitators observe they are cleaner; children wash their hands before using the tablets, and their nails are cut.

5.5 Game-interaction research

The game-interaction research consisted of a questionnaire and a more qualitative evaluation of the game by the children.

5.5.1 *Game-interaction questionnaire*

The game-interaction questionnaire was taken with a stratified sample of the children (247 children). For the questionnaire, optimal scaling techniques were used, to control for assumptions about interval distributions of the children's answers to each question (Gifi 1990; Spss 2001; Meulman & Heiser 1999).
The unidimensionality of the separate scales was evaluated by Principal Component Analysis with Varimax Rotation and by calculating Cronbach's alpha's. Finally, scale scores were obtained by adding optimal scaling quantification scores within scales, and transforming scores linearly to a 0-100 scale. Higher scores indicating more emphasis on the scale's theme.

The overall reliability of the questionnaire was good: Crohnbach's alpha is 0.81. Four different scales could be identified: Usability, Game improves level of knowledge (specific), Game supports learning (in general), and Matching game learning goals (see Table 7).

Table 7 Cronbach's alpha, number of items and example item for the game interaction questionnaire.

Scales	T24	No of items	Example
Usability	.73	6	It was clear what actions I could take in the game
Game Improves level of knowledge	.62	4	I knew how I could become better in the game
Game Supports learning	.57	4	The goal of the game was the same as my own
Matching game learning goals	.58	6	When I wanted to do something in the game I could do that
Total Game evaluation	.81	20	[no example: all items from the Game-interaction research together]

The scale Usability is reliable with a Crohnbach's alpha of 0.73. Given the context in which this questionnaire was answered the reliability of the second scale: Game improves level of knowledge, with a Crohnbach's alpha of 0.62, is also acceptable (George & Mallery, 2003). The other scales are not reliable and will not be used in further analyses.

The average score for Usability was 1.9 on a 5-point scale. This is not a very high score. The average score for Game improves level of knowledge was 2.3 on a 5-point scale.

5.5.2 *Qualitative evaluation game*

All the children indicated they liked the game. When asked what they liked about the game some said it was nice and easy, others thought it was fun and had nice colours. One child mentioned that the game 'had a bit of difficulty'.

The children were asked which mini-game they liked best. Fourteen children mentioned they liked all the mini-games. Other children named one, two or three games they liked very much (see table 8, below). For an overview of all the mini-games used in the game see Appendix G.

Table 8 Number of children that like mini-game.

Mini-game	Frequency	Short description
4. Egg game	41	[Numbers] The children hear a number. They have to drag the right number of eggs into the box.
7. Order shirts	26	[Number discrimination] There are three shirts with numbers on them on a clothesline. Children have to drag the shirts (and the numbers) in the right order.
9. Addition	25	[Addition] The children have to answer an addition by clicking on the number that answers the addition.
14. Three in a row	23	[Addition] An answer to an addition is given. Children have to click the addition that gives this answer. The additions are presented in a matrix. The goal is to have three addition in a row (horizontal, vertical or diagonal).
8. Bus game	21	[Word problem, addition] There is a drawing of a bus with a number of people in it. Then some extra people get into the bus. The children have to say how many people are in the bus now. (Also supported by audio.)
2. Number line	21	[Number discrimination] Children have to put a given number in the right place on a number line.
6. Bead game	16	[Numbers] There is a string of beads. Children have to separate a given number of beads from the other beads on the string.
13. Missing number	10	[Missing number] Three numbers are given. There is a fourth, empty space somewhere between these numbers. The children have to click on the number that should go into this empty space.
1. Number game	8	[Numbers] A number is said (audio) and children have to click on the right amount of objects, or on the right number.
5. Divide in circles	4	[Splitting numbers] Children have to divide a certain amount of object over two circles.
3. Finger game	3	[Numbers] A picture of a hand showing a certain amount of fingers is shown for a limited time. Then the picture disappears and the children have to click on the number that equals the amount of fingers they have just seen.
16. Bigger/smaller	3	[Bigger/smaller] The children have to place two numbers on a number line. Then they have to indicate whether one number is smaller or bigger than the other one.

In addition to this free response to the question which mini-game they liked best, children were asked to rate all the mini-games. Per mini-game children were asked to respond to the following statements:

1 The mini-game works.
2 I understand what I have to do in this mini-game.
3 I like this mini-game.

5.5.2.1 *Mini-game ratings*

In general, the average answers to this questions were between 2.5 and 2.9 on a 5-point Likert scale, which is just above the mean of the scale. For some questions there were no differences between boys and girls. For the other questions, sometimes boys gave a higher score, other times girls gave a higher score. In general the scores are higher for older children, with the exception of 8-year old girls: they almost always gave the highest scores. Table 9, below, gives an overview of the scores given per mini-game, including the differences between boys and girls and between age-groups.

Table 9 Scores per question per mini-game; average and differences for gender and age.

Mini-game		The mini-game works	I understand what I have to do	I like the mini-game
1	Average	2.5	2.6	2.6
	Gender	G>B	G>B	B=G
	Age	9>7	9>7	9>7
2	Average	2.7	2.8	2.7
	Gender	G>B	B=G	G>B
	Age	-	B: 9>7	B: 9>7
3	Average	2.7	2.6	2.7
	Gender	B=G	B>G	B>G
	Age	B: 9>7	B: 9>7	-
4	Average	2.7	2.6	2.7
	Gender	G>B	B=G	G>B
	Age	9>7	9>7	9>7
5	Average	2.6	2.7	2.6
	Gender	G>B	B>G	B>G
	Age	B: 9>7	B: 9>7	B: 9>7
6	Average	2.6	2.7	2.7
	Gender	B=G	B=G	B>G
	Age	-	-	-
7	Average	2.8	2.7	2.7
	Gender	B>G	B>G	B>G
	Age	-	B: 9>7	B: 9>7
8	Average	2.7	2.8	2.7
	Gender	G>B	G>B	B=G
	Age	B: 9>7	B: 9>7	B: 9>7
9	Average	2.8	2.8	2.7
	Gender	G>B	B>G	G>B
	Age	9>7	9>7	-
10	Average	2.5	2.6	2.6
	Gender	G>B	B=G	G>B
	Age	8-year olds give highest scores on all questions		
11	Average	2.7	2.8	2.7
	Gender	B=G	B>G	G>B
	Age	9>7	9>7	9>7
13	Average	2.6	2.6	2.7
	Gender	G>B	B>G	B=G
	Age	-	B: 9>7	B: 9>7
14	Average	2.5	2.6	2.6
	Gender	B>G	B>G	G>B
	Age	B: 7>9	B: 9>7	B: 9>7
		G: 9>7	G: 7>9	-
16	Average	2.8	2.8	2.9
	Gender	B=G	B>G	B>G
	Age	B: 9>7	B: 9>7	B: 9>7

5.5.2.2 *The mini-game works*

Mini-games 2, 3, 9, and 16 received the highest score (2.8) for the question if the mini-game worked. Mini-games 1,10 and 14 were given the lowest score for this question: 2.5. In general, girls gave higher scores on this question than boys; for only two mini-games (7 and 14), boys gave a higher score. Four mini-games were scored equally by boys and girls (3, 6, 11, 16). In general, older children gave higher scores than younger children. As the 8-year old girls gave the highest scores, this was not always the case for the girls: the 9 year-old girls mostly gave a lower score than the 8 year-olds. Also, for mini-game 14 the 7 year-old boys gave the highest score.

5.5.2.3 *I understand what I have to do in this mini-game*

Mini-games 2, 4, 10, 13, and 14 received the lowest score (2.6) on the question if children understood what they had to do. The highest score given for this question was 2.8, for mini-games 2, 3, 8, 9, 11, and 16. The differences in scores are very small, and are only a little higher than the mean of the scale. In general, boys gave a higher score on this question than girls. For seven mini-games they gave a higher score, and for six mini-games there was no differences for scores between boys and girls. Girls gave a higher score only for mini-game 8.

5.5.2.4 *I like this mini-game*

Mini-game 16 was liked best, with an average score of 2.9. Mini-games 1, 10 and 14 were liked least and received the lowest score: 2.6. Four mini-games were liked by boys and girls (1, 3, 8, 13), six mini-games were liked better by girls than by boys (2, 4, 9, 10, 11, 14). The other four mini-games were liked better by boys (5, 6, 7, 16). The differences between scores were very small. These results correspond with the list of mini-games most often mentioned by children when answering the open question on which mini-game they liked best. There are some differences, though: e.g. the bus mini-game, which was mentioned specifically by 21 children answering the open question in what mini-games they liked, received a relatively lower score on the closed question (2.7).

5.5.2.5 *Conclusion*

Mini-game 16 received the highest score on all three questions, whereas mini-games 10 and 14 received the lowest scores on all three questions.

5.5.3 *Visions*

The children were also asked whether they liked the visions; the ‘jobs’ in the overall narrative of the game. Three of the visions were about a boy (goat herder, brick maker and growing vegetables), the other two visions were about a girl (cooking lady, tea lady). For screenshots of the visions used in this pilot, see Appendix H. The average scores for all visions varied between 2.6 and 2.9 on a 5-point Likert scale. Four visions were liked better by girls than by boys, boys only liked the brick maker better. Again 8-year old girls gave the highest scores, except for the vision Growing vegetables. For that vision, 9-year old girls gave the highest score. Older boys gave higher scores for the Goat herder, Cooking lady, Brick maker and the Growing vegetables.

Figure 31 Playing together.

5.6 What factors contributed most to the learning outcomes

Differences between groups on the main dependent variables, were analysed with multivariate analysis of variance for general linear modelling (Multivariate_GLM). The relationship between demographic variables with the main dependent variables was investigated in two steps: in step one Pearson correlations were calculated between demographic variables and research factors. Demographic variables that had at least one significant correlation with a factor were selected. It appeared that all variables had to be selected: Community, Age, Gender, Number of siblings, Row, Family, Mother, Father, Distance to primary school, Distance to secondary school. In step two, the selected demographic variables were added as covariates in all consecutive multivariate-GLM tests to correct for initial differences between participants.

The influence of studied factors on the post-test scores on Math A and B, was tested using two hierarchical regression analyses with Math A or B as independent variables. Results were summarized with the estimated marginal means and the standard errors. Overall, results of statistical analyses were reported as significant with a P value of ≤ .05 or as trends with a P value between .05 and .10. In addition, effect size estimates are evaluated when available, using the partial R squared (R^2), which assists in interpretation of its practical importance. A partial R^2 of. 01 is defined as a small effect size, .06 as a medium, and .14 as a large effect size (Cohen, 1987).

Table 10, below, show the summary of the hierarchical regression analysis of variables predicting final maths scores.

Table 10 Summary of hierarchical regression analysis of variables predicting final mathematics scores.

Variables	Test A-POST		Test B-POST	
	Beta	Delta R^2	Beta	Delta R^2
Step 1: math results		0.21*		0.40*
A-PRE	0.40*		0.12*	
B-PRE	n.a.		0.44*	
Step 2: Demographic variables, baseline		0.06*		0.05*
Age	0.00		0.01	
No of siblings	0.07		-0.03	
Place in row	-0.08#		-0.03	
Has a mother	-0.02		0.00	
Has a father	-0.01		-0.11*	
Mother's education	-0.10*		-0.10*	
Father's education	0.00		-0.01	
Distance to nearest primary school (Km)	0.09*		-0.07	
Distance to nearest secondary school (Km)	-0.09#		-0.04	
Gender	0.02		0.05	
Step 3: Psychosocial factors, baseline		0.02*		0.01*
Self-esteem	0.13*		0.13*	
Self-efficacy	0.11*		0.07*	
Step 4: Logged data		0.07*		0.07*
Percentage complete	0.29*		0.31*	
No of days played/week	-0.07		-0.04	
Step 5: Psychosocial factors at T24		n.a		0.01
Self-esteem	n.a.		-0.08*	
Self-efficacy	n.a.		0.03	
R^2		0.59		0.72
Adjusted R^2		0.35		0.52

*significant P≤.05, # trend p≤.10, n.a.=not applicable because time of measurement is later than A-Post.

As can be seen in Table 10, above, the model with selected predictors explains 59% of the A Post-test and 72% of the B Post-test. Test A-Post scores are mainly defined by the scores at A-Pre (large effect). Test B scores are mainly defined by scores on A-Post and B-Pre test scores (large effect). This could be an indication that children who already have mathematical knowledge before the start of the pilot, benefit most from playing the game. However, as shown in Figures 32 and 33, children significantly improve their scores between pre- and post-tests.

The children were divided in three groups with equal number of children per group: A-Pre scores 0-13; A-Pre scores 14-23; A-Pre scores 24-60. Analysis shows that in test A the children with the lowest scores on the pre-test, improved most (F=29.17, hyp df=2 error df=488, p<.00, eta=.11). At the same time, their absolute score on A-Post was still lower than children with a higher score on A-Pre.

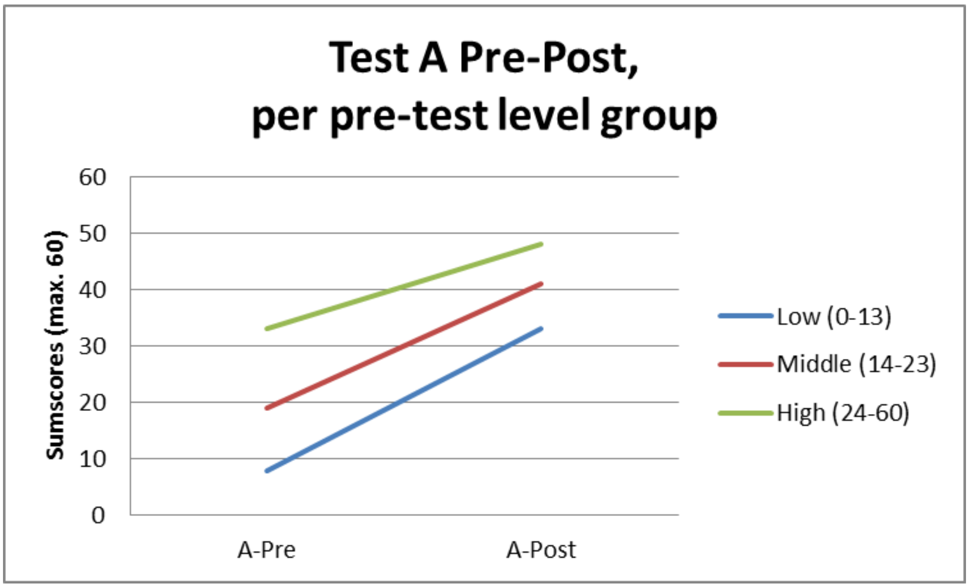


Figure 32 Test A Pre-Post, per pre-test level group.

In test B, this effect has almost disappeared (see Figure 33). On the whole, according to multivariate test there was a trend but not a significant difference between groups (F= 2.48, hyp df=2, error df= 443, *p*=.08, eta= .01). A univariate test contrasting results between groups, however, did differ significantly (*p*=<.00) This indicates that there was still some difference between groups with respect to improvements on test B but not as much as in test A.

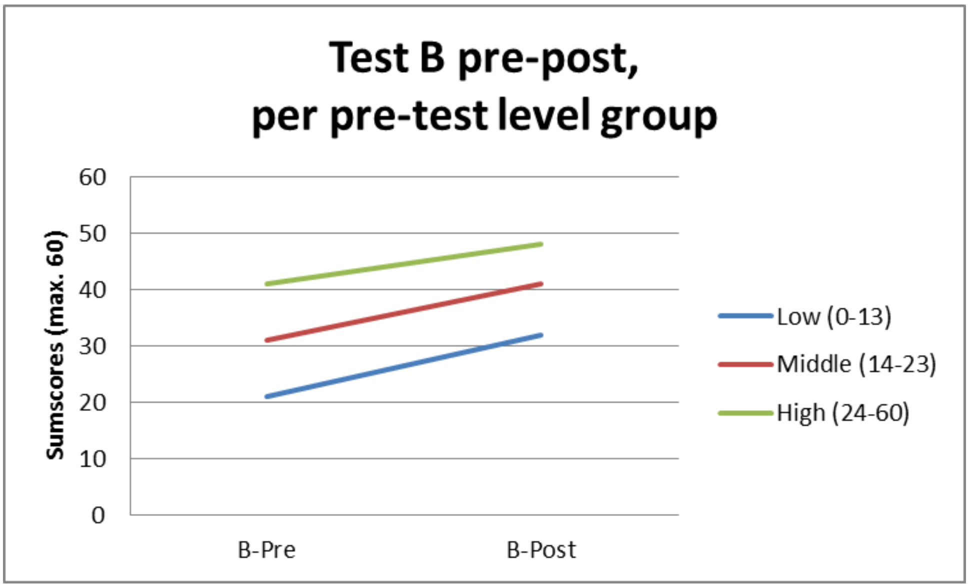


Figure 33 Test B pre-post, per pre-test level group.

Furthermore, demographic variables, Step 2, add a little, but still significantly to the explained variance of the model. The relevant variables have a small to medium significantly effect on the maths results. The higher the education of the mother, the lower the results on both mathematics tests (medium effect). Having a father has a small negative effect on the B-Post. The further the distance to the nearest primary school, the higher the results on A-Post (medium effect).

Step 3, Psychosocial factors in the baseline, also add to the explained variance of the model. A higher self-esteem and self-efficacy at the beginning of the pilot, relate to higher maths scores at the end (medium effect).

The logged data (step 4) can be seen as predictors as well. Specifically percentage complete relates positively to the mathematics scores at the end (medium effect for A-Post; large effect for B-Post).

Finally, Psychosocial factors measured at the end of the pilot (step 5) relate to test B results. Self-esteem has a medium negative relation with test B. However, it is not clear if these scores can be seen as real predictors, as the psychosocial data at the end of the pilot were collected at about the same time as test B-Post was taken. Therefore, causality cannot be assumed.

6 Discussion and conclusion

After a successful first pilot with a mathematics game on a laptop, this study aims to test if children in rural areas in Sudan can learn mathematics autonomously, for a longer period of time, learning diverse and more difficult mathematical concepts. Pilot II lasted 6 months, with 591 children playing the tablet game 45 minutes a day for a maximum of five days a week. The results of 517 children could be used for further analyses. In this chapter the results and the conclusions are discussed. The last paragraph summarises the final conclusions in relation to the research questions.

6.1 Research challenges

Research in remote areas in Sudan proved to be a challenge. One of the reasons for this is that the international research team could not travel to the remote communities themselves. As a result, local researchers had to collect all data using a research protocol. They were trained to use this protocol.
Furthermore, it took ten hours or more to reach some of the communities from Khartoum. Some communities could not be reached in the rainy season. This made logistics hard for the local researchers. They could not always travel at the designated times.

Test protocol
Although facilitators and observers had been trained to use the test protocol, tests were not (always) taken at the designated times. For instance, the pre-test of test B was supposed to be taken at the same time as the post-test of test A. In reality, this was not the case. As a result the pre-test of test B was taken too late, leading to rather high scores on the pre-test. This makes it harder to show significant improvement when comparing to the post-test.

Data collection
Collecting the data themselves took much time, with researchers travelling to the communities to do the testing. Results were written down on paper, and later entered into an Excel file. This was time-consuming, and allowed for human error at the different stages of data entry. Local researchers double checked for data entry mistakes, the research coordinators took random samples and checked paper data with Excel files. This made the data entry even more time intensive, but improved the reliability. As a result, data analyses could only start months after the pilot had finished. If data collection is digitalized, this will facilitate data collection and at the same time reduce human error.

Control group
Using a control group is also an issue: there is an ethical element in asking communities to participate in a pilot as a control group without allowing them to benefit from it. In addition, agreements have to be signed at various levels before communities can participate. As agreement could not be reached in all cases and there was a shortage of potential control groups available, the eventual control groups were smaller than the experimental group. As with the experimental group, the tests were not taken at the designated times in the control groups. This led to the exclusion of the Gedaref control group (pre-test was taken after the children had

started learning), and modified analyses for the North Kordofan (post-test was taken after 3 months instead of 8 weeks) and White Nile (post-test was taken after 6 months instead of 8 weeks) control groups.

Unique child numbers
The accounts in the game should have been based on child numbers. In Pilot I there had been quite a few issues regarding the matching of child data because the children's names were used instead of the child numbers. Arabic names written down in English are not always spelled in the same way (phonetic spelling). Moreover, the children usually have three to four names. In the baseline study, the complete names were used, in the accounts in the game, sometimes only two or three names were used, which made it harder to match the names. Although this was emphasised, in Pilot II the issues were even greater than in Pilot I: in Pilot II 146 account in the game used the names of the children, in Arabic. 386 accounts were entered with the name written in English. Due to this, only 508 of the 532 log files could be matched to children. To ensure this does not happen again, an obligatory field was added to the form used to create new accounts. It is now not possible to create a new account without including a unique child number.

Logged data
The game uses an online – offline system. This means children can always play the game on their tablet, at their own level; their progress can be synchronized to a central server whenever there is internet connectivity. In reality, the internet was unreliable, and logged data were downloaded manually from the tablets to laptops, and then on a USB stick that was brought to Khartoum by hand. Although this was a time-consuming procedure, it was the only way to collect logged data in this pilot. Due to the time involved, logged data were only collected at the end of the pilot. As a result, information on the actions of children became available at the end of the pilot and could not be used to feed back into the programme during the pilot.

The management system in the game also allows facilitators to upload tablet data to a laptop locally. The local management system then gives insight in the attendance, actions and progress of the children in one community. This feature could have been used during this pilot, but was not. In the future, facilitators should be more aware of this option and trained to use it to allow them to track progress of the children in their own group during the pilot.

Timestamp
As the tablets were not online, there was a problem with the timestamp for about 20% of the tablets. The date on some tablets was not correct, e.g. 1970. This, in itself, is not a problem for the management system, if the date is not changed during the pilot. However, for some tablets this date was changed to 2015 after an update of the game was installed. The number of months played, as registered in the management system, was therefore not correct, and could not be used for further analyses. Having an internet connection would solve this problem: all tablets would automatically have the right date. In the meantime, facilitators should make sure tablets have the right timestamp at the beginning of the pilot.

6.2 Drop out

A total of 62 children (10%) are considered as drop out during the pilot.
The definition of drop out used was: Children who stopped playing the game before the pilot ended AND had not finished 90% or more of the game. Children who stopped playing the game before the pilot ended in March 2015, but had finished the game (90% or more) were not considered to be drop outs. Neither were children who stopped playing for a period of time, but resumed playing the game before the pilot ended. Although a 10% drop-out is substantial, this is less than drop-out in regular education, which can reach up to 50% (The World Bank, 2012; Sriprakash 2010), and similar to a tablet intervention for mathematics in a school in Malawi (Pitchford, 2015). Only 5% of this drop out (3 children) was reported as caused by children or parents refusing the programme. The other 95% (59 children) were reported by facilitators to have dropped out because of their families moving (to find water or harvest crops).

This drop out was significantly related to state; White Nile had a significant higher percentage of drop outs than the other two states. Age and gender did not influence drop out, which indicates that boys and girls and different age groups have a similar risk of dropping out. This is a positive finding, showing that the learning solution does not put children at a disadvantage based on gender or age. Children with less siblings dropped out significantly more often than children with more siblings.
The reason for this is not entirely clear.

The applied gaming solution with tablets allows children to learn mathematics without a teacher. This is, therefore, a flexible learning solution. Children could take the tablet with them, when they travel to another community, and keep on learning. In this pilot, however, we chose to carry out rigorous research. This meant that children had to stay in one village to stay involved. All children that moved to another village were consequently seen as drop outs. When the programme is scaled up, more communities will be involved. This increases the possibility of children moving to another community that is using the game as well. Based on the logged data, a child can continue playing the game at his or her own level. Allowing children to take a tablet or enrolling them in the same programme in another community would probably reduce drop out significantly.

6.3 Learning outcomes

6.3.1 Learning effects in general

The results show that the game is effective in significantly increasing the scores on the mathematics tests A and B. The average increase of scores on test A was larger than for test B. Although there was some overlap between test A and test B, this does not explain the notably high scores on the pre-test of test B. This is probably due to the fact that the pre-test of test B was taken too late. Children had continued playing the game in the meantime, and thus increased their mathematical knowledge.

To absolutely ensure that the increase in scores is due to playing the game, a control group was included. The control group was enrolled in informal education in out-of-school centres. Unfortunately, tests A pre and post were not taken at the right time with the control group, which limits the possibilities to compare between the experimental and the control group. For North Kordofan and White Nile, the control group data could be used. Analyses show that there are no significant differences between the results of the experimental group and the control group in North Kordofan, while the control group had three times as much learning time (two lessons of 45 minutes a day, instead of one, and a three-month interval between pre and post-test, instead of eight weeks). For the White Nile control group, there were small, significant differences with the experimental group: the control group had a greater increase in scores. The White Nile control group had six times as much learning time (two lessons of 45 minutes a day and a six month interval between pre and post-test). Although it would have been better to have the control group tested at the designated times, these results are still positive. The children in the experimental condition reach (almost) the same results as the children in the control groups, although they have had significantly less learning time.

Only 57% of the children completed the game (90% or more), 77% of the children completed 50% or more of the game. There are four possible explanations for this: we have seen that most children played for an average of 2-3 days a week. It is, therefore, possible that they did not play enough days to finish the game. Looking at the learning outcomes, we can conclude that children have learned much. Even when compared to regular education for a much longer period of time. This suggests that their progress is faster than children in regular education. Another explanation may be that the part of the curriculum that was developed, takes more time to finish than planned. As this was the first time this part of the game was tested with children, we did not know how much time they would need to finish. To ensure they could keep on learning during the pilot, rather many different mini-games were included. It is possible that only the children with more informal knowledge, or the children who learned very quickly, could finish the game. Thirdly, the game may have been too difficult for the children, leading to a slower pace of learning. Again, looking at the learning results, this does not seem to be the case. Finally, the children may not have liked the game, resulting in less active engagement. However, the results of the game interaction questionnaire show that children did like the game, and that scores for the different mini-games varied between 2.5 and 2.9 on a 5-point scale, which is between average and slightly above average.

The results of EGMA testing, carried out by independent consultants, came back very positive. Compared to children in Khartoum and Jordan who had been to school for 2 years and 5 months, the children in the pilot on average have the highest scores in three subtests of EGMA. In a fourth subtest, the children in the pilot had on average a higher score than the children in Khartoum. In three other subtests the children in the pilot only had slightly lower scores than the children in Khartoum and Jordan, although they had only played the game for about five months. In the last three subtests, the children in the pilot had much lower scores than the children in Khartoum and Jordan; the reason for this was that these subjects were not taught in the pilot. As EGMA is an internationally validated test and because it was carried out by independent consultants, this is external, independent evidence that children have learned from playing the game.

6.3.2 *Differences in learning effects for gender, age or state*

There were no significant differences in learning outcomes for gender, which is a very positive finding; boys as well as girl learned by playing the game.

There were significant differences for age, with the younger children having a lower score than the older children. The reason for this is probably that older children have gained more informal knowledge before the pilot. The differences for age were larger in test A and decreased in test B. This can probably be explained by the fact that the younger children learned more by playing the game and caught up with the older children.

There were significant differences for state, with White Nile having a higher score than North Kordofan and Gedaref on test A, pre and post. The can be explained by the higher average age in White Nile: older children have higher test scores. At the same time, North Kordofan, with a lower pre-test score on test A, shows the largest increase in scores. The average age in North Kordofan is the lowest, predicting lower scores on the pre-test of test A. Although the average score on the post-test is not as high as the average score in White Nile, children in North Kordofan have, on average, increased their scores most. For test B, Gedaref differs significantly from the other two states, showing a smaller learning effect.

6.3.3 *Correlations between the three methods*

There are significant positive correlations between the scores on the mathematics tests and EGMA, and between the mathematics tests and the percentage complete (logged data). Although all tests show significant positive correlations with EGMA, the B-Post-test has the strongest correlations. This can be explained by the fact that test B measures more difficult mathematical concepts than test A, and is, therefore, more similar to EGMA than test A. These correlations show that test A and B measure similar concepts of mathematics as EGMA.

This means that the tests and the logged data measure similar mathematical concepts as EGMA. As EGMA is an internationally validated test, this adds to the validity of the tests A and B and the logged data. In addition, these significant, positive correlations support the conclusion that children have learned. If three different types of measuring instruments show similar results, conclusions based on these results are stronger and more reliable. Finally, this information can be used to reduce the number of tests in the test protocol when scaling up in the same context. The rigorous research carried out in this pilot, was rather time-consuming.
When this game is introduced in a similar context, we could rely on a combination of percentage complete and EGMA (after about 1 year, and again after 2 years).
This would significantly reduce the time spent on testing and data entry. In a new context, or with a new subject, the test protocol should be as rigorous as used in this pilot.

Although we had intended children to play the game five times a week during the pilot period, only about 6% of the children managed to do so. Most of the children (about 50%) played 2-3 days a week on average. Some played 2-3 days every week, others played for two weeks and then skipped a week. It is a negative finding that we did not manage to involve the children for five days a week, especially because a learning session only takes 45 minutes. At the same time, it is a very positive finding that children have achieved the significant mathematics results, described above, in only half of the time allowed. Finally, these results show that the game allows children to be absent for some time, and then resume the programme successfully. This flexible way of learning is difficult to achieve in regular education: if children miss out on instruction for a week or more, it becomes very hard to catch up and benefit from further instruction.

Logged data showed that children who made less mistakes in the game (more mini-games finished without any mistakes) had higher scores on the mathematics tests and EGMA. This was unrelated to how much of the game they completed; children making less mistake did not move faster through the game. There are two possible explanations for this: either children have more mathematical knowledge, which also shows in the tests, or these children work more accurately and make less mistakes that could have been avoided. It would be interesting to know if the playful environment of the game in any way stimulates children to answer (too) quickly. They might bring this strategy to the test.

6.4 Psychosocial research

6.4.1 Effect on psychosocial factors that influence learning

It has proved to be difficult to measure the psychosocial factors that are known to influence learning. The overall reliability of the psychosocial questionnaire was good (Crohnbach's alpha= 0.71). But, although the questionnaire was based on existing scales and questionnaires, not all scales were reliable enough to use for further analysis. This can probably be explained by the fact that using questionnaires with young children, and especially in developing countries, is a challenge. These children have probably never thought about themselves in this way. It is worthwhile to try and improve the psychosocial questionnaire, though, because a reliable questionnaire that is easy to answer has a considerable added value.

Factor analysis indicated five scales could be identified. These differed slightly form the a-priori scales. Two of these scales: Self-efficacy and Motivation & future orientation were reliable enough to include in further analyses. Self-esteem was not included as a scale, but was measured using four different trees representing the level of self-esteem. Self-esteem is the only factor that showed a significant increase in scores between the two measurements at T=0 and T=26 (from 1.9 to 2.5 on a 5-point Likert scale). Being involved in the pilot and playing the game has had a positive effect on children's self-esteem. This is consistent with research on self-esteem and learning: they mutually reinforce each other, although increasing self-esteem by itself does not automatically increase academic results *(*Baumeister, Campbell, Krueger & Vohs, 2003). There were no significant differences between boys and girls. This is a positive finding; playing the game does not affect boys and girls in a different way. Average scores for all scales were around 50%, which does not provide much guidance for improvement.

6.4.2 Other psychosocial effects

The results from the focus group meetings were very positive. Children, parents and facilitators liked the game and appreciated the programme. Although these results are different from the psychosocial questionnaire, where scores were around 50%, they did provide extra insight in additional benefits of the programme. Children reported they have made friends with the other children in their learning sessions, and now play with each other. Parents reported they like the fact their children are safe and taken care of during the learning sessions. Facilitators report that children's hands are cleaner, and their nails are cut, because of using the tablets.

6.5 Game interaction research

The overall reliability of the game interaction questionnaire was good: Crohnbach's alpha is 0.81. However, although the questionnaire was based on an existing game interaction questionnaire, not all the scales were reliable enough to include in further analyses. Factor analysis indicated that four scales could be identified. These scales were slightly different from the a-priori scales. Only two of these scales: Usability and Game improves level of knowledge were reliable enough. The average score for Usability was 1.9 on a 5-point scale, the average score for Game improves level of knowledge was 2.3 on a 5-point scale.

The below-average score of Usability might be an indication that children found it difficult to use the game. There can be several reason for this: children did not know how to use a tablet, children did not understand what actions they could take in the game, mini-games did not work properly or children had difficulty to understand what to do with a mini-game. However, observations showed that children learned how to use the tablet in about two hours. The fact that children have played an enormous number of mini-games and watched specific instruction videos multiple times, suggests that they understood what actions they could take in the game. This would suggest that children had no overall difficulty in using the tablet and the game.

During the pilot there were two games that did not function properly: the Egg game and the Bead game. The bugs in these games were repaired and afterwards the children could play these games without any problems. When asked which mini-game they liked best, most children reported they liked the egg game best.
The bead game came in as seventh best liked mini-game. In addition, to the question if the mini-games worked, the Egg game and the Bead game both received above average scores (2.7 vs. 2.6 on a 5-point scale). This suggests that this does not explain the below average score of Usability. When asked if they knew what to do with a mini-game, average scores per mini-game varied between 2.6 and 2.8 on a 5-point scale. This is also an above average score, and does not explain the below average score of Usability.

Per mini-game the children were asked three questions: Does it work?, Do you understand what you have to do?, and Do you like this mini-game?. Per question, girls gave higher scores for some mini-games, and boys gave higher scores for other mini-games. For some mini-games, girls and boys gave the same scores.
In general, the 8-year old girls gave the highest scores. This suggests that in the variety of mini-games offered, children will always find a few mini-games they really like. This may contribute to the fact that there are no differences in learning outcomes between boys and girls.

In general, girls gave higher scores for the visions than boys. Boys only liked the brick maker better than girls. The reason for this might be that boys are more competition focused, and use the stars (shown per vision) as an indication of their progress. This is similar to findings in the formative evaluation of the game with Arabic speaking children of the Weekend school in The Hague (Stubbé, Badri, Telford, Oosterbeek & van der Hulst, in press).

6.6 What factors contribute most to learning outcomes

Analyses show that the scores on the pre-tests contribute most to scores on the post-tests. This might imply that playing the game does not support learning mathematical knowledge, but that informal knowledge, or intelligence in general, before the pilot started determines the learning outcomes. Looking at the increase of scores on the tests shows that this is probably not the case. All children significantly improve their knowledge on both tests, and in test A children who have the lowest scores on the pre-test show the highest increase of scores. Although statistical regression (scores on a post-test tend to show a lower standard deviations that the score on the pre-test) probably influences this as well, there is a significant difference between the groups with the lowest pre-test score and the highest pre-test score. Percentage complete in the game also contributes to learning outcomes: the higher the percentage complete, the higher the score on the tests. This is to be expected as percentage complete is a measurement of mathematical skills in itself.

Furthermore, demographic variables contribute a little, but still significantly to the learning outcomes. For instance, a higher education of the mother and having a father relate to lower maths results. It is not clear how this can be explained.
The longer the distance to the nearest primary school, the better the learning outcomes. This might be explained by the fact that children are more motivated to learn because they have no access at all to regular education.

It is remarkable that gender and age do not significantly explain variance of the model. In educational research, gender and age usually have an effect on learning outcomes, with differences in learning styles (Severiens & ten Dam,1994) and older children having acquired more informal knowledge. That we did not find that in this study can be explained by the fact that the game supports learning of girls as well as boys. Both boys and girls like different parts of the game and mini-games.
But perhaps we should look in a different direction as well: classroom dynamics are completely different in game-based learning. For instance, research shows that in classrooms in Kenya teachers engage more boys than girls in the learning process (Limboro, 2015). In addition, some Kenya teachers humiliate girls in class through derogatory terms. Game-based learning is absolutely neutral, boys and girls are actively engaged in the same way.

A higher self-esteem and self-efficacy at the beginning of the pilot, relate to higher mathematics scores at the end. This means that children who think they can learn mathematics and who believe in themselves have higher scores on the post-test. There can be two explanations for this: either children have a higher level of mathematical knowledge, and are therefore more confident they can learn.

Or children who believe they can learn mathematics have higher scores on the post-tests. Even without knowing the cause and effect of this relation, it is a positive finding that self-esteem has increased during the pilot, as it has a large effect on learning outcomes.

The game interaction factors do not contribute to learning outcomes. The scores on Usability and Game improves level of knowledge do not relate to the learning outcomes. This is somewhat remarkable. We might expect lower scores on the mathematics tests if children feel the usability of the game is bad, or if they feel the game does not help them to learn mathematics. An explanation for this might be that these factors contribute indirectly to the learning outcomes, through motivation to learn. Unfortunately, factor analysis did not support the a-priori scale Motivation to learn. The scale Future orientation and Motivation came up as a combined scale. Due to this, Motivation to learn could not be used as a separate scale in the regression analysis.

6.7 Conclusion

In this paragraph, the findings per research questions, described in the introduction of this report, are summarised.

1 Do children learn mathematics by playing the game for a longer period of time (sustained learning)?

   This pilot has shown that all children improved their mathematical abilities by playing the game. The fact that there is a correlation between total amount of time (months) played and the post-test scores also shows that the children continue to learn when they play the game for a longer period of time. There were no significant differences for gender. The differences between age groups, with older children performing (slightly) better than younger children, is to be expected and follows the normal development of children. There are significant differences between states, with White Nile performing best on test A, and Gedaref having the lowest scores on test B. Children in North Kordofan show the largest increase in scores between pre-test and post-test. There are significant positive correlations between the three methods used to measure learning effects (mathematics tests, logged data, and EGMA). This means that all three methods show similar results, which makes the overall conclusions about learning effect stronger. These findings also indicate that in the future it is not necessary to use all three methods, when the game is used in the same or a similar context.

2 What are the (psychosocial) effects of learning with technology on children and the communities they live in?

   To answer this research question a questionnaire, focus group discussions and interviews were used. Unfortunately, only three constructs of the questionnaire were reliable: Self-esteem, Self-efficacy and Motivation and future perspectives. Self-esteem has significantly increased during the pilot. Self-efficacy and Motivation and future perspectives have not changed significantly. This means children have gotten a better opinion about themselves, either because they have learned mathematics or because of other factors in relation to the pilot:

e.g. social aspects of learning together, visits to the community by others, or the use of ICT. In addition, both Self-esteem and Self-efficacy show a significant positive effect on learning outcomes. The focus group discussions and interviews show positive effects as well. The children like to learn, they like the game, have started playing together more, show better behavior, and keep their hands cleaner. Parents appreciated the fact that they knew their children were well taken care of during the learning sessions. Now children have started learning, they would like to continue, and learn more subjects as well.

3 What do children think of the game?

All the children liked the game. They thought it was fun and liked the colours. One child said it had a bit of difficulty. Unfortunately, only two construct of the questionnaire about game interaction were reliable: Usability (1.9 on 5-point scale) and Game improves level of knowledge (2.3 on a 5-point scale).
The average scores on these construct were (slightly) below average, but showed no significant correlations with learning results. The qualitative evaluation showed that both boys and girls liked the game, but they appreciated different elements (mini-games and visions). This may explain why boys and girls show similar learning results; although their preferences may differ, they could find elements in the game they liked.
The scores on: ‘The mini-game works’, and ‘I know what to do in this mini-game’ were average with very little variation (2.5 to 2.8 on a 5-point scale).

4 What factors contribute most to learning effects?

A multivariate regression analysis on all the factors that were reliable, and/or showed significant change during the pilot, showed that the scores on the pre-tests contributed most to learning effects. This does not mean that the knowledge children have before starting the game determines the results. Children with a lower score on the pre-test had a significantly higher increase of scores than children with a higher score on the pre-test. Although ceiling effect and statistical regression might influence this, this shows that children learn from the game. This is supported by the fact that percentage complete (logged data) contributes to learning outcomes: the higher the percentage of the game completed, the higher the score on the post-tests. As mentioned before, Self-esteem and Self-efficacy have a positive effect on learning outcomes.
The longer the distance to the nearest primary school, the higher the learning effect. Children might be more motivated if there is no real alternative to access education. The level of education of the mother and having a father both have a negative effect on learning outcomes. At the moment, it is not clear why.
Gender and age do not relate to learning outcomes, which is a positive finding: boys as well as girls, from different age-groups, can learn from the game.

5 What are the characteristics of the children who dropped out?

As drop out is always an issue in pilots that run for a longer period of time, specific attention was paid to this. A total of 62 children dropped out, 10% of the children. According to the remarks made by facilitators, the reason for 5% of the drop out was that either parents of children ‘refused the programme’. The other 95% of the drop was, according to facilitator remarks, caused by families moving to another community to find water or harvest the crops. Analyses have shown no significant differences for gender or age. Boys as well as girls, of all age-groups, have dropped out. Although this is a positive finding, as it shows that the intervention does not exclude specific groups, it does not provide any information that could be used to decrease drop out. Children in White Nile had a significant higher risk of dropping out. The explanation for this is that in two communities in White Nile almost all children have dropped out. Facilitators reported that many families in these communities had moved during the pilot period.

7 References

Adams, G. R., & Marshall, S. K. (1996). A developmental social psychology of identity: Understanding the person-in-context. *Journal of Adolescence*, *19*, 429–442.

Adel Abdullah Idris Siddiq (2015). Field Survey Report on the E-Learning Sudan Project (ELS). Results of the Analysis of EGMA performance of ELS children.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review*, *84*(2), 191.215.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., Vohs, K. D. (2003). Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles?. *Psychological Science in the Public Interest 4* (1): 1–44. doi:10.1111/1529-1006.01431. ISSN 1529-1006.

Beetham, H., & Sharpe, R. (Eds.). (2007). *Rethinking pedagogy for a digital age: Designing and delivering e-learning*. Abingdon, UK, & New York, NY: Routledge.

Bitew, G. (2008). Using plasma TV broadcasts in Ethiopian secondary schools: a brief survey. *Australian Journal of Educational Technology*, 24, 2 150-167.

Bloom, B. S. (ed.) (1985). *Developing Talent in Young People*. New York: Ballentine Books.

Bodovski, K., and Farkas, G. (2007) Mathematics Growth in Early Elementary School: The roles of Beginning Knowledge. *Student Engagement and Instruction* (p. 115-130).

Burnett, N. (1996). Priorities and strategies for education – a World Bank review: the process and the key messages. *International Journal of Educational Development*, Vol 16, No 3, pp. 215-220.

Carroll, J.B. (1963) A model of school learning. *Teachers College Record, 64*, 723-733.

Chard, D.J., Baker, S.K., Clarke, B., Jungjohann, K., Davis, K. & Smolkowski, K. (2008) Preventing Early Mathematics Difficulties: The Feasibility of a Rigorous Kindergarten Mathematics Curriculum. *Learning Disability Quarterly*, Volume 31, Winter 2008.

Cheek, J. M., Smith, S.M., & Tropp, L. R. (2002, February). Relational identity orientation: A fourth scale for the AIQ. Paper presented at the meeting of the Society for Personality and Social Psychology, Savannah, GA.

Clark, R.E., Yates, K., Early, S. & Moulton, K. (2008). An analysis of the failure of Electronic Media and Discovery-based learning: Evidence for the performance benefits of Guided Training Methods. In Silber, K.H. & Foshay, R. (Eds.) Handbook of Training and Improving Workplace Performance, Volume 1: Instructional Design and Training Delivery. Somerset, NJ: Wiley

Clarke, D. J. (2002). E-learning: Big bang or steady evolution? E-learning Technologies 2002, January 2002.

Cohen (1987). Statistical power analysis for the behavioral sciences. L. Erlbaum Associates

Collis & Moonen (2001). Pedagogy: making the U-turn. In: Flexible learning in a digital world. Open University Netherlands.

Copely, J.V. (Ed.) (1999). *Mathematics in the early years* Reston, Virginia: National Council of Teachers of Mathematics.

Cremin, T., & Arthur, J. (Eds.). (2014). *Learning to teach in the primary school* (3rd ed.). Abingdon, UK, & New York, NY: Routledge.

Cremin, P. & Nakabugo, M. G. (2012). Education, development and poverty reduction: A literature critique. *International Journal of Educational Development*, 32 (2012) 499-506.

Daniel, J. (2010). Mega-Schools, Technology and Teachers: achieving education for all Routlegde: London and New York

Dewey, J. (1938). Experience and Education. Collier Macmillan Publishers, London.

Duncan, G.J., Claessens, A., Huston, A.C., Pagani, L.S., Engel, M., Sexton, H., Dowsett, C.J., Magnuson, K., Klebanov, P., Feinstein, L., Brooks-Gunn, J., Duckworth, K. & Japel, C. (2007) School Readiness and Later Achievement. *Developmental Psychology* 2007, Vol. 43, No. 6, 1428-1446.

Fleuren, M. A. H., Paulussen, T. G. W. M., van Dommelen, P., van Buuren, S., (2012). Towards a measurement instrument for determinants of innovations. *International Journal for Quality in Health Care*, *26*(5), 501-510.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. Simulation & Gaming, 33(4), 441-467.

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.

Gifi, A. (1990). Nonlinear multivariate analysis. Chichester: Wiley.

Ginsberg, H.P., & Russel, R.L. (1981) Social class and racial influences on early mathematical thinking. *Monographs of the Society for Research in Child Development, 46(6)*, 1-69.

Greenman, E., Bodovski, K., and Reed, K. (2011). Neighbourhood characteristics, parental practices and children's math achievements in elementary school. The Pennsylvania State University, PA, United States

Guberman, S.R. (1996). The Development of Everyday Mathematics in Brazilian Children with Limited Formal Education. *Child Development*, Volume 67, Issue 4, pages 1609–1623, August 1996. DOI: 10.1111/j.1467-8624.1996.tb01817.x

International Labour Office (ILO), International Programme on the Elimination of Child Labour (IPEC). Geneva: ILO, 2013.

Jonassen, D.H., Peck, K.L & Wilson, B.G. (1999). Learning with technology: a constructivist perspective. Merril, 1999.

Jones, A. & Naylor, R. (2014). The quantitative impact of armed conflict on education in Nigeria: counting the human and financial costs. CfBT Education Trust.

Jonker, V. & Wijers, M. (2008). Th!nklets for mathematics education. Re-using computer games characteristics in educational software. Paper presented at the International Conference of the Learning Sciences (ICLS). from http://www.fi.uu.nl/en/icls2008/550/paper550.pdf.

Kallaway, P. (2001). The need for attention to the issue of rural education. *International Journal of Educational Development*, 21 (2001) 21-32.

Kanu, A.S., Hengel, C. van (2013). A Report on the Baseline Assessment of Children's Early Literacy Abilities in Selected Primary Schools in Classes I, II, III. Woord en Daad, 2013.

Ke, F. (2008). A Case Study of Computer Gaming for Math: Engaged Learning from Gameplay? *Computers & Education* 51 (2008) 1609-1620.

Kilpatrick, J., Martin, W.G. & Schifter, D. (Eds.) (2003) *A research companion to principles and standards for school mathematics*. Reston, Virginia: National Council of Teachers of Mathematics.

Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2000). Experiential learning theory: Previous research and new directions. In R. J. Sternberg, & L. F. Zhang (Eds.), *Perspectives on cognitive, learning, and thinking styles*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H.W. (1992). Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept. An interim test manual and research monograph. Macarthur, New South Wales, Australia: University of Western Sydney, Faculty of Education.

Lange, A., Evers, A., Jansen, H., & Dolan, C. (2002). PACHIQ-R. The Parent-Child Interaction Questionnaire- Revised. Family Process, 41, 709-722.

Lepper, M. R., & Iyengar, S. S. (1999). Rethinking the role of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*. *76*, p. 349-366.

Li, Q. & Ma, X. (2010). A Meta-Analysis of the Effects of Computer Technology on School Students' Mathematics Learning. *Educational Psychology Review* (2010) DOI: 10.1007/s10648-010-9125-8.

Limboro, C. (2015). Gender and teacher pedagogical practices in schools in Nairobi informal settlements and Kilifi County, Kenya. Proceedings UKFIET, 13th international Conference on Education and Development: Learning for Sustainable Futures. Oxford, 15-17 September 2015.

Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science: A Multidisciplinary Journal*, *54*, 333–369.

Mayes, T., & de Freitas, S. (2007). Learning and e-Learning: The role of theory. In H. Beetham, & R. Sharpe (Eds.), *Rethinking pedagogy in the digital age* (pp. 13-25). Abingdon, UK, & New York, NY: Routledge.

Meulman, J. J., Heiser, W. J., & SPSS Inc. (1999). *Categories 10.0*. Chicago, IL: SPSS Inc.

Milo, B. (2003), Mathematics instruction for special-needs students. Thesis University of Leiden, Heerenveen: Brouwer & Wielsma.

Oprins, E., Visschedijk, G., Bakhuys-Roozeboom, M., Dankbaar, M., Trooster, W. & Schuit, S. (2015). The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method. *International Journal of Technology Enhanced Learning*, Volume 7, Issue 4, DOI: 10.1504/IJTEL.2015.074189.

Pellegrino, J.W., & Goldman, S.R. (1987). Information processing and elementary mathematics. *Journal of Learning Disabilities, 20*, 23-32.

Pintrich, P. R., & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33-40.

Pitchford, N.J. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in Psychology*, April 2015, Vol 6, article 485.

Power, T., Gater, R., Grant, C. & Winters, N. (2014). *Educational Technology Topic Guide.* HEART Topic Guides. London: The Health & Education Advice & Resource Team (HEART). Department for International Development (DFID).

Praet, M. & Desoete, A. (2014). “Enhancing young children’s arithmetic skills through non-intensive computerized kindergarten interventions: A randomized controlled study”. Teaching and Teacher Education 39 (2014) 56-65.

Räsänen. P., Salminen, J., Wilson, A.J., Aunio, P. & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development* (2009), DOI:10.1016/j.cogdev.2009.09.003.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., López, X., López, V., Rodriguez, P. & Salinas, M. (2003). Beyond Nintendo: Design and Assessment of Educational Video Games for First and Second Grade Students. *Computers & Education* 40 (2003) 71-94.

Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54-67.

Schmitt, D. P. & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, *89*, 623-642.

Serafini, T. E., Maitland, S. B, & Adams, G. R. (2006, March). *The Functions of Identity Scale: Revisions, validation and model testing*. Poster presented at the Biennial Meeting of the Society for Research on Adolescence, San Francisco, California.

Severiens, S.E. & ten Dam, G.T.M. (1994). Gender differences in learning styles: a narrative review and quantitative meta-analysis. Higher Education 27: 487-501, 1994.

SPSS (2001). CATPCA. Statistical support – algorithms [SPSS website]. Available at: http://www.spss.com/tech/stat/Algorithms.html; September 12, 2001 [accessed May 5, 2003].

Sriprakash, A. (2010). Child-centered education and the promise of democratic learning: Pedagogic messages in rural Indian primary schools. *International Journal of Educational Development*, Vol. 30, pp. 297-304.

Steenbergen-Hu, S. & Cooper, H. (2013). A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on K-12 Students’ Mathematical Learning. *Journal of Educational Psychology* 2013, Vol. 105, No. 4, 970-987.

Stubbé, H. & Theunissen, N.C.M. (2008). Self-directed adult learning in a ubiquitous learning environment: a meta-review. Proceedings of Special Track on Technology Support for Self-Organised Learners, during 4th EduMedia Conference 2008 "Self-organised learning in the interactive Web" - A change in learning culture? 02. - 03. June 2008 in Salzburg.

Stubbé, H., van der Klauw, M. & Langefeld, J.J. (2014). E-learning Sudan – Pilot II. Descriptive analysis of the baseline study. TNO 2014 R11794.

Stubbé, H., Badri, A., Telford, R., van der Hulst, A.H. (2015). E-Learning Sudan, formal learning for out-of-school children. The 9th European Conference on Games Based Learning ECGBL 2015. Nord-Trondelag University College, Steinkjer, Norway. 8-9 October 2015.

Stubbé, H., Badri, A. Telford, R., Oosterbeek, S., Hulst, A. van der (in press). Formative evaluation of a mathematics game for out-of-school children in Sudan.

Stubbé, H., Badri, A., Telford, R., van der Hulst, A.H. & van Joolingen, W. (in press). E-Learning Sudan, formal learning for out-of-school children.

The World Bank (2012). *The status of the education sector in Sudan.* African Human Development Series, 66608) UNESCO

Timmermans, R. (2005) Addition and subtraction strategies. Assessment and instruction, Nijmegen: Radboud University.

Trinder R. (2013). Business students' beliefs about language learning in a university context. *English for Specific Purposes*, *32*, 1-11.

UNESCO (2013). Toward universal learning what every child should learn (Report No. 1 of 3). Quebec, Canada.

UNICEF (2009). Open and Distance Learning for Basic Education in South Asia.

UNICEF (2015). Education under fire: How conflict in the Middle East is depriving children of their schooling. Amman, Jordan.

Unwin, T. (2009). ICT4D implementation: policies and partnerships. In Unwin, T. ICT4D. Cambridge: Cambridge University Press, 125-176.

USAID (2009) *Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children*. EdData II Technical and Managerial Assistance, Task Number 2. Contract Number EHC-E-02-04-00004-00 Strategic Objective 3. December 31, 2009.

Vogel, 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, 229-243.

Vygotsky, L. (1978). Mind in Society. London: Harvard University Press.

Wouters, P., van der Spek, E.D. & van Oosterndorp, H. (2009). Current Practices in Serious Game Research: A review from a Learning Outcomes Perspective. IGI Global.

Wouters, P., van Nimwegen, C., van Oostendorp, H. & van der Spek, E.D. (2013, February 4). A meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*. Advance online publication. DOI: 10.1037/a0031311.

A Geographic characteristics of the communities

State – name of the state
Locality – name of the locality
Community – name of the community
Characteristics - of community: rural, nomad or IDP
GIS coordinates – of the community
Distance to nearest primary school – from the community
Distance to nearest secondary school – from the community

Every community was assigned a specific code, based on state, locality and community names.

Code	State	Locality	Community	Charac-teristic	GIS/Co-ordinates	primary school	**secondary school**
NK-S-EF	North Kordofan	Shaykan	Elfatih	rural	13°02'30.6"N 30°14'29.4"E	5KM	15KM
NK-S-OK	North Kordofan	Shaykan	Om Kateera	rural	13°10'24.6"N 30°29'26.0"E	7KM	30KM
NK-S-AR	North Kordofan	Shaykan	Alaradeeb	rural		7KM	25KM
NK-B-OG	North Kordofan	Bara	Om Hagar	rural	13°37'45.1"N 30°26'41.3"E	10KM	8KM
NK-B-ER	North Kordofan	Bara	Elrgeba	rural	13°39'40.0"N 30°24'44.1"E	3KM	8KM
NK-E-AG	North Kordofan	Elrahad	Agabtina	rural	12°46'45.3"N 30°40'34.0"E	12KM	12KM
NK-E-RI	North Kordofan	Elrahad	Redina	IDP	12°47'27.8"N 30°38'31.3"E	3KM	12KM
WN-G-MO	White Nile	Gooli	Mona	rural	13°04'26.9"N 32°31'40.4"E	9KM	12KM
WN-G-OT	White Nile	Gooli	Om Tifag	rural	13°02'48.6"N 32°27'22.2"E	12KM	7KM
WN-G-ET	White Nile	Gooli	ElTalha	rural		7KM	13KM
WN-T-ET	White Nile	Tandalty	ElTben	rural	13°00'26.3"N 32°08'24.3"E	7KM	25KM
WN-T-OD	White Nile	Tandalty	Om Dresa	rural	12°55'22.8"N 32°10'09.3"E	7KM	25KM
WN-T-OG	White Nile	Tandalty	Om Gowa	rural	13°06'10.9"N 31°51'09.2"E	14KM	25KM
WN-S-GT	White Nile	Elsalam	GoriEltkeal	rural		9K	35K

Code	State	Locality	Community	Charac-teristic	GIS/Co-ordinates	primary school	**secondary school**
GA-F-ER	Gadaref	Alfawo	ElRimela	IDP/rural	14°06'16.9"N 33°56'36.6"E	6K	30K
GA-F-WN	Gadaref	Alfawo	WadNorien	voyager/rural	14°07'31.7"N 34°10'09.6"E	5K	12K
GA-F-ED	Gadaref	Almafaza	ElDabseen	voyager/rural		8K	17K
GA-G-TH	Gadaref	Galabat	TaybaLahweenA	voyager/rural	13°48'19.4"N 35°13'55.4"E	6K	29K
GA-G-TA	Gadaref	Galabat	TaybaLahween B	voyager/rural	13°48'19.4"N 35°13'55.4"E	6K	29K
GA-W-WD	Gadaref	Wasat AlGadaref	WadSanad	IDP/rural	13°59'08.1"N 35°20'47.9"E	7K	9K

B Questionnaire demographic characteristics children

1. Unique child number for pilot
2. Child’s name
3. State
4. Community
5. Age
6. Gender
7. Number of siblings in the family
8. Place of this child in row of siblings (1=oldest)
9. Father and mother in family
10. Highest level of education of mother
11. Highest level of education father

C Mathematics tests A, B, C and D

Test A

1. Can you count to 10?

2. Can you count on from 4?

3. Can you count on from 6?

4. Can you count back from 5?

5. Can you count back from 7?

6. Can you tell me how many tomatoes you see?
   [picture with 3 tomatoes]

7. Can you tell me how many tomatoes you see?
   [picture with 1 tomato]

8. Can you tell me how many tomatoes you see?
   [picture with 2 tomatoes]

9. Can you point at the picture which has the most carrots
   [two pictures with: 1 carrot and 3 carrots]

10. Can you point at the picture which has the most carrots
    [two pictures with: 2 carrots and 5 carrots]

11. Can you say what this number is?
    3

12. Can you say what this number is?
    6

13. Can you say what this number is?
    2

14. Can you say what this number is?
    8

15. How many carrots do you see? Point at the right number.
    [picture with 3 carrots]
    [1 3 5]

How many tomatoes are this? Point at the right number.
[picture with 1 tomato]
[1 2 6]

16. Can you write down the number 3?

17. Can you write down the number 12?

18. Can you write down the number 9?

19. I have one tomato, and I buy two more. How many tomatoes do I have?

20. There are two people in the bus, and three more people get in the bus. How many people are there in the bus?

21. What number comes after the number 7?

22. What number comes before number 6?

23. Can you point at the highest number
    5 – 16 – 10

24. Can you point at the lowest number?
    11 – 17 – 9

25. Which number should be in the empty box?

9	10		12	13

26. Can you say the answer to this sum?
    3 + 1 =

27. Can you say the answer to this sum?
    2 + 1 =

28. Can you say the answer to this sum?
    1 + 4 =

29. Can you say the answer to this sum?
    3 + 2 =

30. Can you say the answer to this sum?
    1 + 1 =

Test B

1. Can you count to 20?

2. Can you count on from 8 [to 20]?

3. Can you count back from 16?

4. How many tomatoes do you see? [picture with 8 tomatoes]

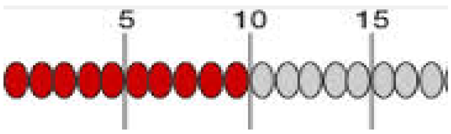
5. Can you point at the picture with most carrots?
[two pictures with: 12 carrots and 16 carrots]

6. Can you say what number this is? [17]

7. Can you write down the number 8?

8. Can you write down the number 14?

9. Point at the bead that is number 12 in the line.



10. Point at the person who is number 16 in the queue.
[picture with a queue of 20 people, in 4 groups of 5 people]

11. Can you what number should be in the empty box?

8		10	11	12	13

12. Can you point at the picture which has most goats?
[four pictures with: 16 goats, 9 goats, 19 goats, and 15 goats]

13. Point at the number that is the lowest.
[13 11 18]

14. Point at the number that is the highest.
[18 15 9]

15. Point at the picture with the least tomatoes.
[three pictures with: 6 tomatoes; 3 tomatoes; 4 tomatoes]

16. Point at the number that is bigger than 18.
[12 16 20]

17. Point at the number that is smaller than 15?
[12 16 19]

18. Can you put the numbers in the right order (from low to high)?
[13 -14 -12 -11 – 15]

19. Can you put the pictures in the right order (from most people to least people)?
[four pictures with: 6 people; 12 people; 9 people; 15 people]

20. Can you count back from 20 in steps of 2?
[20, 18 ….]

21. Can you fill in the empty boxes with the right numbers?

2	4	6		

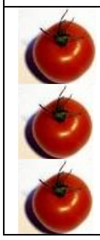
22. Fatma has 4 carrots. She gets 2 more. How many carrots does she have now?

23. There are 7 children in one hut. There are 2 children in another hut. How many children are there in total?

24. Can you say the answer to this sum?
[4+1=]

25. Can you say the answer to this sum?
3+ 5 =

26. How many tomatoes do you need to make 6?

6	
	

27. Can you split number 8 in three different ways?

8	

28. Can you say the numbers you need to make 10? (horizontal)

10	
7	
	5
3	
	4

29. Can you say the answer to this sum?
5 + 2 =

30. Can you say the answer to this sum?
6 + 2 =

Test C

1. Which number is missing in the question mark?

10		?		14	15

2. Which number is bigger than 14?
[12 16 13]

3. What is the answer to this sum?
3+ 5 = ?

4. Can you split number 12 in two different ways?

12	

5. There are 10 people in a bus. 6 more people get on the bus. How many people are in the bus now?

6. There are 13 people in a bus. 3 more people get on the bus. How many people are in the bus now?

7. Can you say the right answer to this sum?
12 + 2

8. Can you say the right answer to this sum?
15 + 3

9. How much is 11 + 1?
Can you point out the picture with the right amount of sheep?
[Three pictures with: 10, 12, and 13 sheep]

10. Your neighbour has 15 tomatoes, he buys 2 more. How many tomatoes does your neighbor have now?

11. You have two strings of beads, see below.
How many beads do you have?

12. There are 5 people in a bus. In the next village, more people get in the bus.
There are 8 people in the bus now.
How many persons got on the bus in the next village?

13. Can you say the right answer to this sum?
8 + 5

14. Can you say the right answer to this sum?
7 + 7

15. Which sum leads to number 15?
9 +5
9 + 6
13 + 3
16 + 2

16. There are 8 people in a bus. 3 more people get in the bus.
How many people are in the bus now?

17. There are 9 people in a bus. 6 more people get in the bus.
How many people are in the bus now?

18. How many eggs have these two egg boxes together?

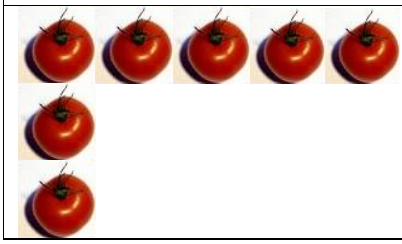
 

19. Can you point at the picture with the right amount of goats?
8+9
[Three pictures of: 12 goats, 15 goats, 17 goats]

20. Which number should be in the empty space?
6 + 12 = □

21. Which number should be in the empty space?
15 + □ = 20

22. How many tomatoes do you need to fill in the right side to have 14 tomatoes?

<table>
<tr><td colspan="2">14</td></tr>
<tr><td></td><td></td></tr>
</table>

23. How many eggplants do you need to fill in the left side to have 17 eggplants?

17	
	

24. How much is 5 +0?

25. How much is 8 + 0?

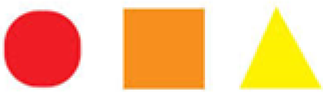
26. There are 12 people in a bus. In the next village nobody gets in the bus. How many people are in the bus now?

27. Can you point at the picture with the right amount of goats?
   6-2
   [three picture of goats: 4 goats, 6 goats, 8 goats]

28. You have 9 carrots. You give 3 carrots to your neighbour.
How many carrots do you have now?

29. There are 6 people in a bus. 3 people get out of the bus in the next village.
How many people are in the bus now?

30. Which shape is a triangle?



Test D

1. There are 12 people in a bus. In the next village 5 more people get in the bus.
How many people are in the bus now?

2. Can you say the right answer to this sum?
4 + 7 =

3. Can you say the right answer to this sum?
9 + 4 =

4. How would you split number 12 into two similar numbers?

12	

5. Can you split the number 14 in 2 different ways?

14	

6. There are 5 people in a bus. In the next village nobody gets in the bus.
How many people are in the bus now?

7. Can you say the right answer to this subtraction?
2 – 1

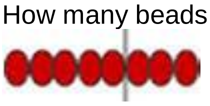
8. Can you say the right answer to this subtraction?
7 – 3

9. Can you point at the picture with the right amount of carrots?
8 – 2
[three pictures of carrots: 6 carrots, 8 carrots, 10 carrots]

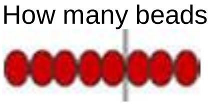
10. There are 9 people in a bus. In the next village 4 people leave the bus.
How many people are in the bus now?

11. Can you say the right answer to this subtraction?
9 – 5 =

12. You have a string of beads (see below). You lose 3 beads.
How many beads do you have now?



13. You have a string of beads (see below). You lose 5 beads.
How many beads do you have now?



14. You have 6 eggs. You drop the eggs, and some are broken. You have 3 eggs left.
How many eggs were broken?

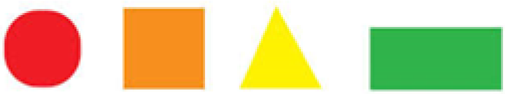
15. There are 12 people in a bus. In the next village 4 people leave the bus.
How many people are in the bus now?

16. Which number should be in the empty space?
5 - □ = 2

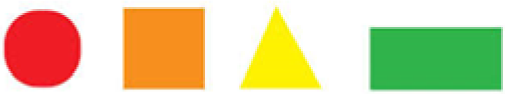
17. Which number should be in the empty space?
9 - □ = 7

18. Can you say the answer to this subtraction?
2 – 2 =

19. Which shape is a rectangle?

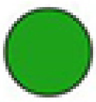
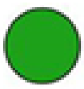
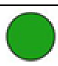
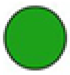


20. Which shape is a triangle?

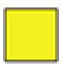
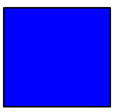
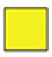
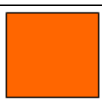
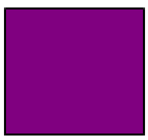


21. This ball is like a shape. Which shape looks like this ball?

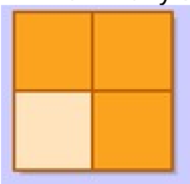
22. Which shape is the smallest?

a	b	c	d	e
				

23. Which shape is the biggest?

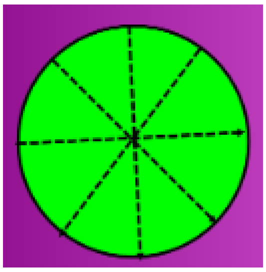
a	b	c	d	e
				

24. How many squares do you see?



25. In this flag you can see two different types of shapes, could you name them?

26. In this figure there are two different types of shapes, could you name them?



27. There are 15 people in a bus. In the next village 4 people get out of the bus. How many people are in the bus now?

28. Can you say the right answer to this subtraction?
17 – 2 =

29. Can you say the right answer to this subtraction?
15 – 7 =

30. How many sides does a rectangle have?

D Psychological questionnaire

Self-esteem	Ask the children to pick the tree that matches their feelings about themselves.		
Self-efficacy	1	I am certain I can accomplish my goals	Pintrich & De Groot (1990)
	2	I can handle whatever comes my way	Pintrich & De Groot (1990)
	3	I stay confident, even when things are difficult	Pintrich & De Groot (1990)
	4	I can do most things as well as my friends	Self-description Questionnaire II, Marsh (1992)
	5	I expect to do well in my learning	Self-description Questionnaire II, Marsh (1992)
Motivation	6	I like learning new things	Self-directed learning, (Stubbé & Theunissen, 2009)
	7	I think learning new things is important	Self-directed learning, (Stubbé & Theunissen, 2009)
	8	I am doing my best to learn new things	Self-directed learning, (Stubbé & Theunissen, 2009)
Future orientation	9	I know what I want to become	The Functions of Identity Scale (Serafini, Maitland & Adams, 2006)
	10	I know what I want to be	The Functions of Identity Scale (Serafini, Maitland & Adams, 2006)
	11	In the future, I will do as well as my friends or better	Psychosocial Vulnerability Assessment Tool (ILO, 2013)
Social support	12	When I have a problem, I can talk to my parents or family	Fleuren, Paulussen, van Dommelen & van Buuren (2012). Lange, Evers, Jansen & Dolan (2002).
	13	I feel supported to learn by my parents or family	Fleuren, Paulussen, van Dommelen & van Buuren (2012).
	14	I feel supported to learn by others in my community	Fleuren, Paulussen, van Dommelen & van Buuren (2012).
	15	I am accepted by my community	Psychosocial Vulnerability Assessment Tool (ILO, 2013)
	16	I feel like I am part of the group	Psychosocial Vulnerability Assessment Tool (ILO,

			2013)
	17	I participate in activities in my community	Psychosocial Vulnerability Assessment Tool (ILO, 2013)
	18	I feel I can trust others in my community	Psychosocial Vulnerability Assessment Tool (ILO, 2013)
Identity orientation	19	My feeling of being a unique person, being distinct from others is important to me	Personal Identity, AIQ IV (Cheek, Smith & Tropp, 2002)
	20	My reputation, what others think of me, is important to me	Relational identity, AIQ IV (Cheek, Smith & Tropp, 2002)
	21	My relationships with people I feel close to is important to me	Social identity, AIQ IV (Cheek, Smith & Tropp, 2002)
	22	My feelings of belonging to my community is important to me	Collective identity AIQ IV (Cheek, Smith & Tropp, 2002)

E Game interaction questionnaire

Open question, write down answer

1. What did you like about the game?

Questions 2-21 are closed questions, use the cups for a 5-point scale

Feedback

2. I received enough instruction and support in the game
3. I knew how well I did in the game, all the time
4. While playing the game, I felt I was getting better
5. I knew how I could become better in the game

Challenge

6. The content of the game was too difficult for me
7. The content of the game was too easy for me
8. I could play different mini-games to learn what I had to learn
9. The difficulty of the game matched my own level

Control

10. When I wanted to do something in the game, I could do that
11. The game allowed me to do the things I wanted to do
12. The game made it hard to do what I wanted to do (R)
13. I could experience myself what worked in the game and what did not work

Rules & goals

14. Before I started playing the game I knew what the goals of the game were
15. It was clear how I could reach the goals of the game
16. I knew what goals I wanted to reach in the game
17. The goal of the game was the same as my own

Action language

18. It was easy to learn how to play the game
19. The game worked in the way I expected it to work
20. It was clear what actions I could take in the game
21. I found it easy to use the game

F Characteristics of children with missing data, per instrument

Missing test A

		A Pre	A Post
Gender	Boy	2	3
	Girl	0	10
Age	7	0	5
	8	0	3
	9	2	5
State	White Nile	1	2
	North Kordofan	0	10
	Gedaref	1	1

Missing test B

		B Pre	B Post
Gender	Boy	6	27
	Girl	10	20
Age	7	8	28
	8	4	13
	9	4	6
State	White Nile	2	2
	North Kordofan	10	33
	Gedaref	4	12

Missing Logged data

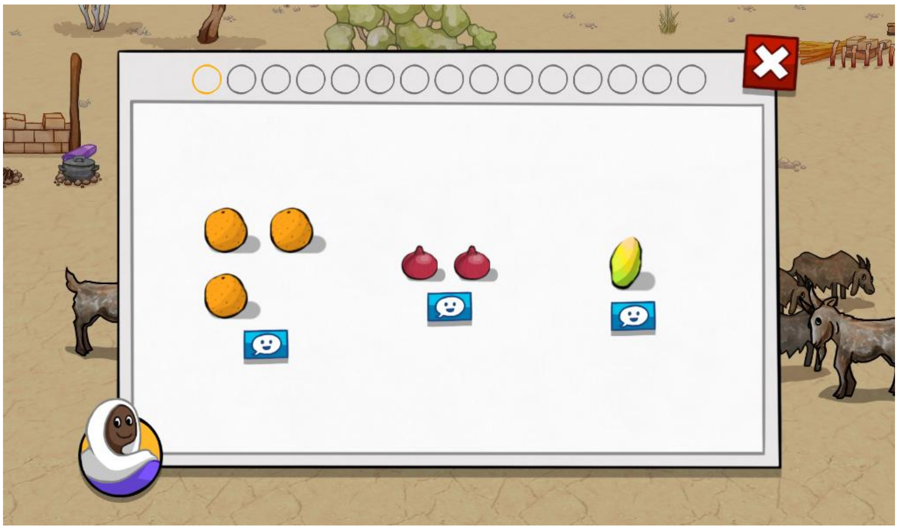
		Logged data
Gender	Boy	45
	Girl	32
Age	7	35
	8	17
	9	25
State	White Nile	15
	North Kordofan	29
	Gedaref	33

Missing Psychosocial questionnaire

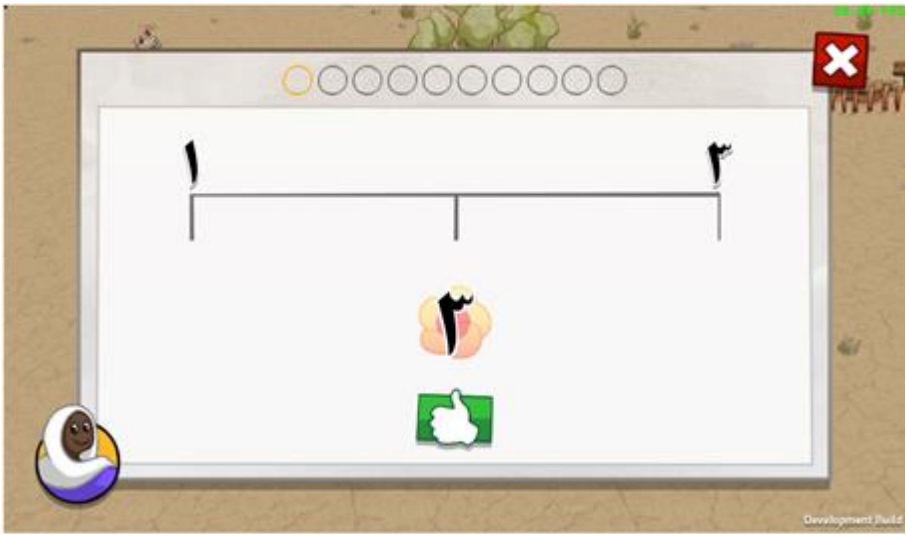
		PS T0	PS T26
Boy/Girl	Boy	2	14
	Girl	0	12
Age	7	0	6
	8	0	9
	9	2	11
State	White Nile	1	4
	North Kordofan	0	0
	Gedaref	1	22

G Mini-games in the game

Mini-game 1: Numbers, 1-10
[Numbers]
The child hears a number. Then the child has to click on the picture with the right amount of objects.



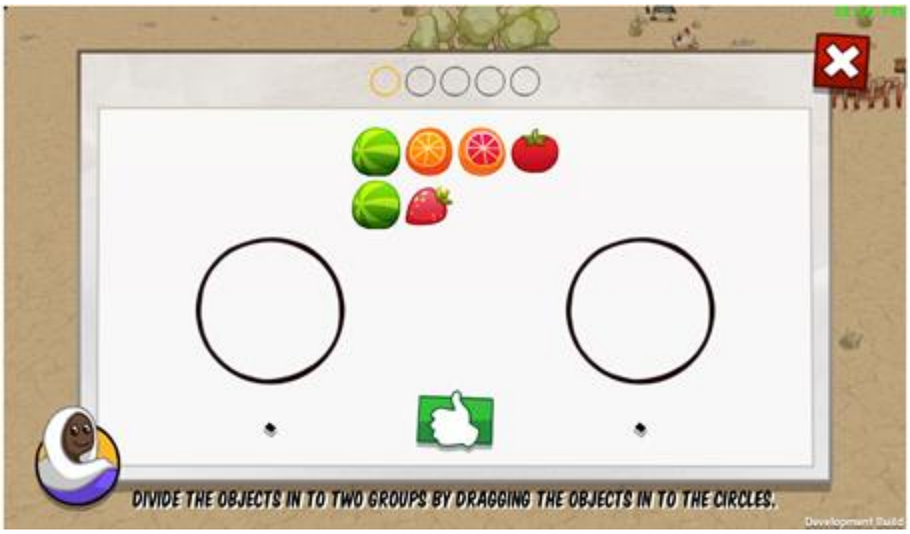
Mini-game 2: Numberline
[Number discrimination]
A number is given, the child has to place this number in the right position on the number line.



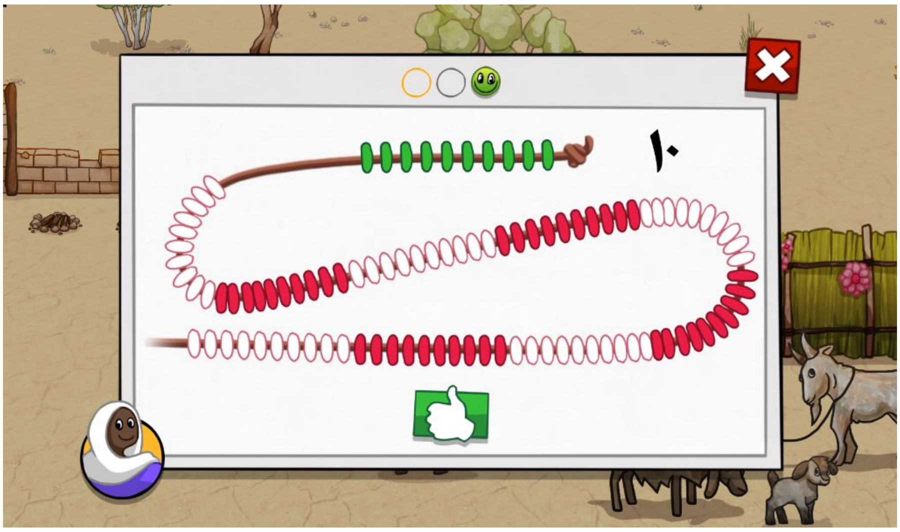
Mini-game 3: Finger game
[Numbers]
A picture of a hand showing a certain amount of fingers is shown for a limited period of time. Then the picture disappears and the children have to click on the number that equals the amount of fingers they have just seen.

Mini-game 4: Egg game
[Numbers]
The children hear a number. They have to drag the right number of eggs into the egg box.

Mini-game 5: Divide in circles
[Splitting numbers]
Children have to divide a certain amount of objects over two circles.

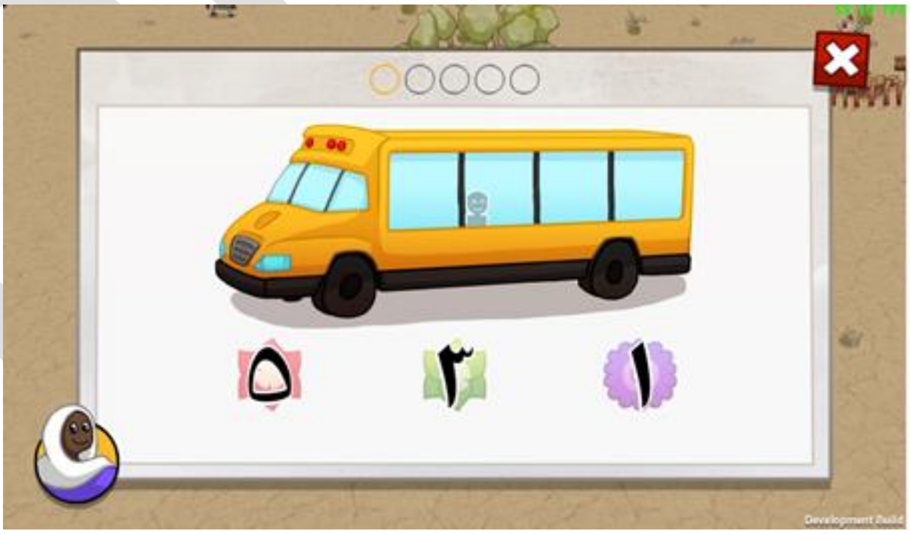


Mini-game 6: Bead game
[Numbers]
There is a string of beads. Children have to separate a given number of beads from the rest of the beads on the string.

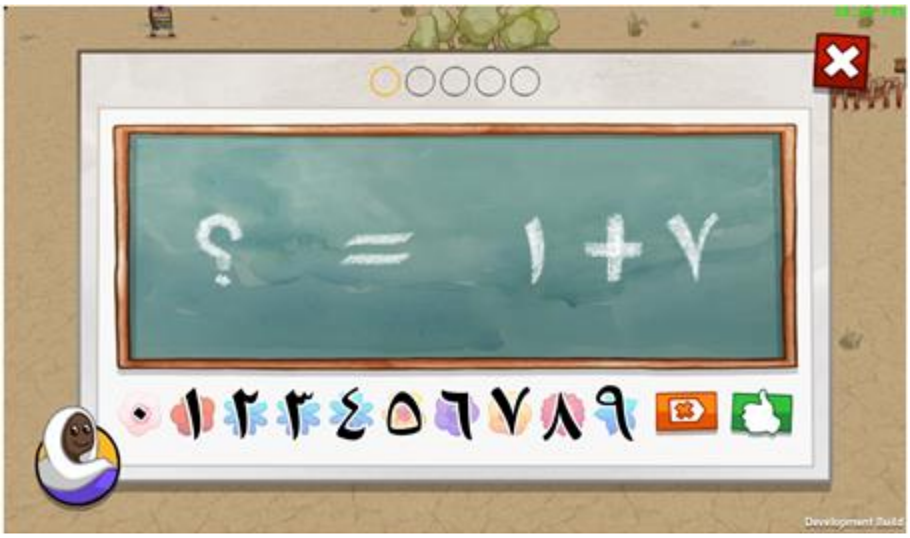


Mini-game 7: Order shirts
[Number discrimination]
There are three shirts with numbers on them on a clothesline. Children have to drag the shirts (and the numbers) into the right order.

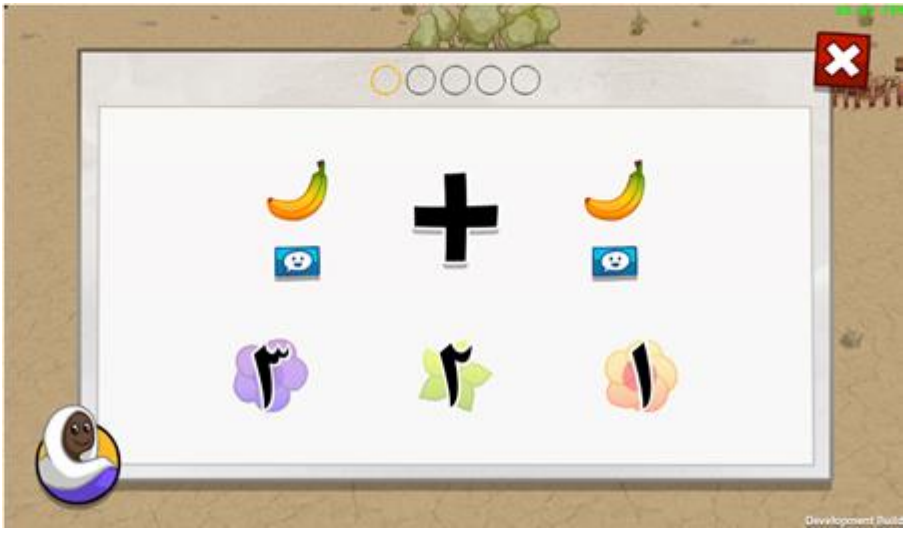
Mini-game 8: Bus game
[Word problems, addition]
There is a drawing of a bus with a number of people in it. Then more people get in the bus. The children have to click on the right number of people that is in the bus now (also supported by audio).



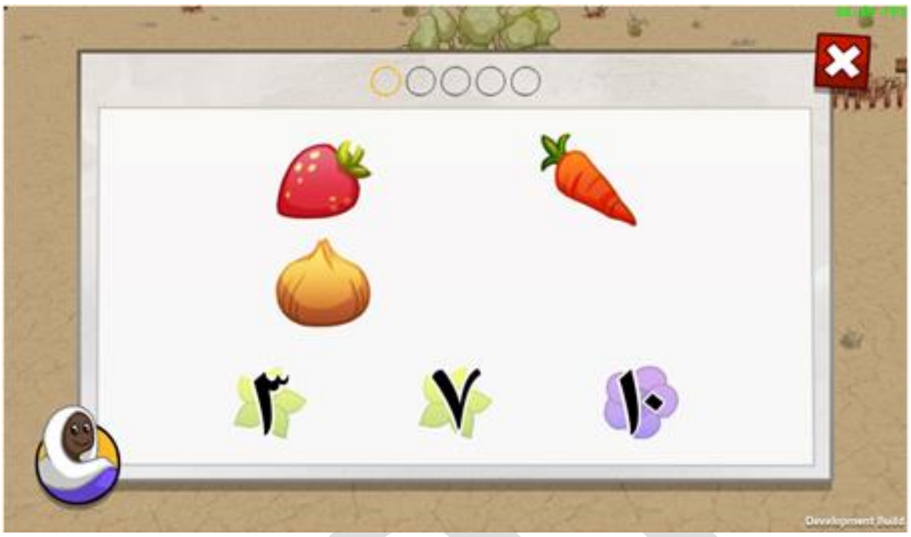
Mini-game 9: Addition problems
[Addition]
Children have to answer an addition by clicking on the number that answers the addition.



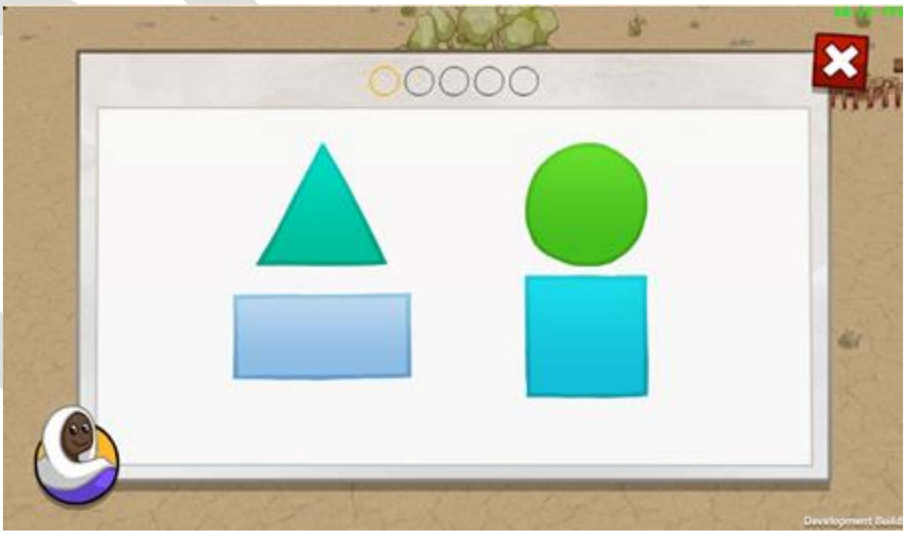
Mini-game 10: Addition with objects
[Addition]
An addition is shown, with objects instead of numbers. The child has to answer the addition by clicking on the number that equals the answer.



Mini-game 11: Numbers, 1-20
[Numbers]
Children have to click on the number that is the same as the amount of objects shown.



Mini-game 12: Shapes, recognition
[shapes]
Different shapes are shown. The child has to click on the shape that is mentioned in the audio.

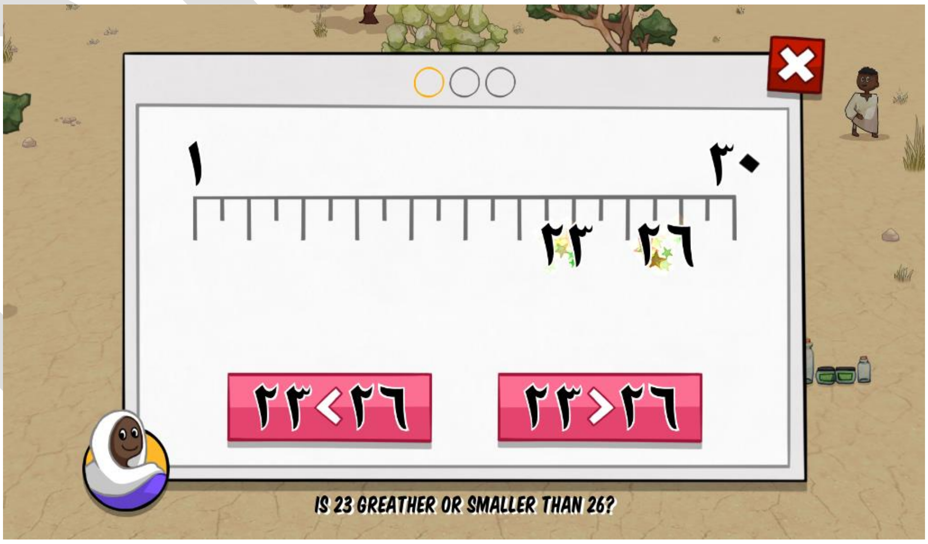


Mini-game 13: Missing number
[Missing number]
Three numbers are given. There is a fourth, empty space somewhere between these numbers. Children have to click on the number that should go into this empty space.

Mini-game 14: Three in a row
[Addition]
An answer to an addition is given. Children have to click on the addition that gives this answer. The additions are presented in a matrix. The goal is to have three additions in a row (horizontal, vertical or diagonal).

Mini-game 15: Find the shapes
[Shapes]
A picture with many shapes is shown. Children have to find as many shapes of a certain type as they can.

Mini-game 16: Bigger or smaller
[Bigger/smaller]
The children have to place two numbers on a number line. Then they have to indicate whether one number is smaller or bigger than the other one.



H Visions in the game

The visions shown below are the ones used in the 6 month pilot.

Vision 1: Goat herder

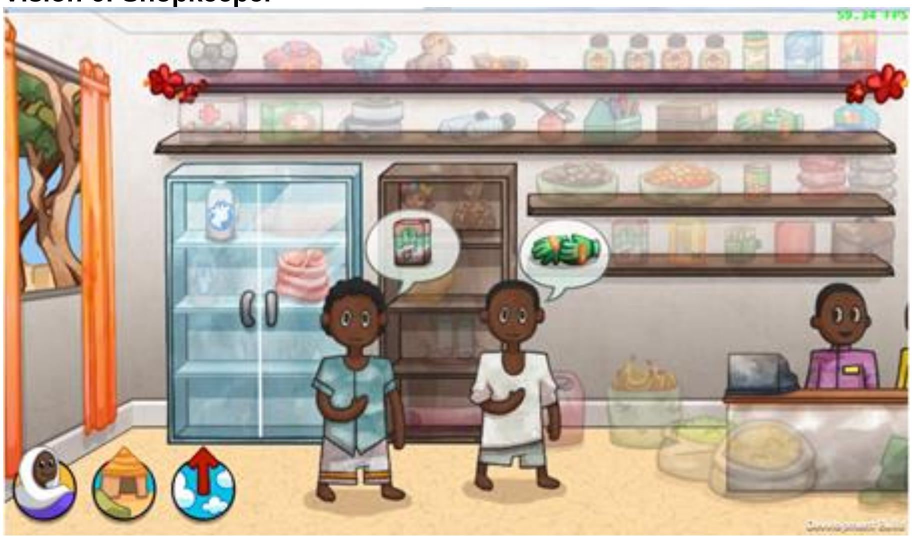
Vision 2: Cooking lady

Vision 3: Brick maker

Vision 4: Growing vegetables, girl

Vision 5: Growing vegetables, boy

Vision 6: Shopkeeper



Distribution list

The following agencies/people will receive a complete copy of the report.

2 ex. WarChild
Innovation Programme Manager
Attn. K. Radford
Helmholtzstraat 61G
1098 LE Amsterdam

PDF TNO:
H.E. Stubbé
M. van der Klauw
J.J. Langefeld
N.C.M. Theunissen
A.H. van der Hulst
W.S.M. Piek (research manager)

1 ex. Archive, TNO location Soesterberg