The SWAN game-based approach to learning foundational number language: a feasibility study

Main Public Report

Chris Donlan, Carolyn Bruce, Vanessa Meitanis, Caroline Newton

University College London

Acknowledgements

The authors would like to thank all those who participated in the study, and the families and professional teams who gave such strong support. We would also like to thank our long-standing advisory group for their generous and expert guidance.

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this report, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

Nuffield Foundation

1. Executive Summary

A. Background.

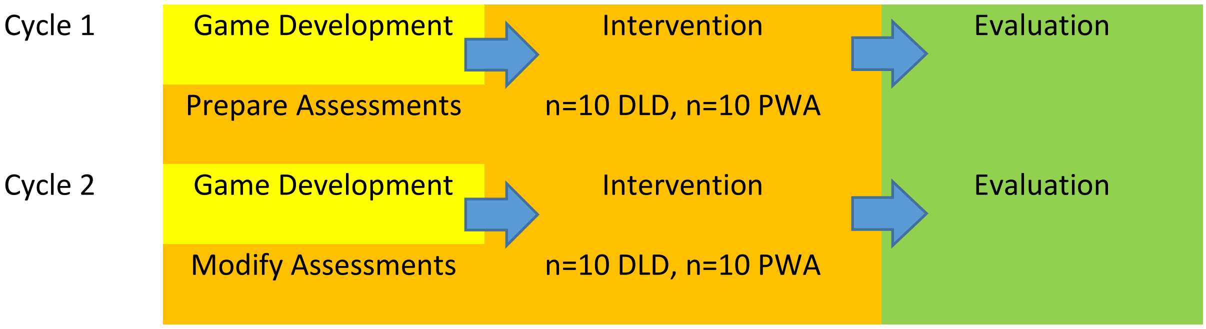
Around 8% of children in the UK have Developmental Language Disorders (DLD). Several research studies have shown that most children with DLD have difficulty acquiring basic numeracy skills. Yet Donlan et al. (2005) found that children with DLD had relatively good underlying number concepts, suggesting that, if foundational skills could be firmly established, enhanced mathematical development might be expected. In addition, a combination of published research and clinical experience suggest that a high proportion of people with aphasia (PWA) are likely to have substantial difficulty with number processing, and that this may be more prevalent than in the general population. The commonality of effects on basic numeracy in children with DLD and adults with aphasia, is striking. However, these are areas of research and practice which are pursued, for the most part, independently. This project sought to examine the underlying constraints on basic numeracy in each group through the development and evaluation of an evidence-based intervention with applicability across both client groups.

A prototype of the Sequences in Words and Numbers (SWAN) game was developed by researchers at University College London, in partnership with SoftV, a software company specialising in computer games. It sought to use the motivational power of 'gamification' to provide repeated and concentrated input of the number sequence for children and adults with language difficulties whose basic number knowledge was limited.

Initial pilot studies indicated that significant modifications to the game would be required to accommodate individual differences in number knowledge and rates of learning. The current feasibility study therefore sought (a) to implement these changes through collaboration with SoftV, and (b) to evaluate the possible utility of the SWAN game as an intervention to support basic number knowledge through a series of single case intervention studies carried out with samples of children with DLD and adults with aphasia.

B. Study Design.

A single case design was used, based on repeated assessment of basic number knowledge before and after three weeks of daily gameplay sessions, each session lasting around 15 minutes. The study was planned to involve two Cycles, with evaluation of results from Cycle 1 guiding modifications to the game and the assessments for Cycle 2.



C. Cycle 1: Effects of COVID-19.

The study commenced in October 2019 with major modifications to the game. The initial levels of the game were simplified in terms of number range and progression. The 70 levels of the prototype were expanded to 140 levels representing increasing sequence complexity. The basic structure of each level, and the algorithm which regulates players' progress through the game, remained unaltered from the prototype. Each level presents three successive 'boards' to the player. On each board, the player must touch adjacent tiles in consecutive order, and touch the final tile again to clear the sequence. Scores are based on the length of sequences entered. Progress through the levels was dependent on achieving a specified score.

The final board from Level 1 of the revised version of the game is shown below.

A 'board' from Level 1, a 3x3 matrix, showing score accumulated from previous boards and 'help' arrows which appear when players are slow to initiate a sequence. Scores, based on sequences completed across all three boards at each level, must reach criterion in order to proceed to the next level.

Board size varies through levels, according to the range and complexity of the sequences introduced. The following table gives an indication of the ways in which aspects of the game are manipulated through the progression from Level 1 to Level 140.

Level	Board Size	Sequence Type	Number Range
1	3x3	Ascending	1-3
20	6x6	Ascending	13-20
50	5x5	Descending	7-1
100	7x7	Ascending Even Numbers	2-52
140	4x4	Ascending in 5s	5-20

Cycle 1 was underway when the lockdown of 23rd March 2020 closed schools and prevented PWA from travelling to assessment appointments. The status of Cycle 1 at lockdown was as follows:

PWA (Target 10 participants):

Complete baseline measurements on six participants

Gameplay data on six participants

Partial outcome measurements on six participants

Children with DLD (Target 10 participants):

Complete baseline measurements on seven participants

Less than 30% gameplay data on seven participants

No outcome measurements

The effects of the pandemic required modification of the study design, and a complete revision of its implementation. An expanded sample for cycle 2 (n=15 per group, instead of n=10) was sought, and the entire procedure of assessment, distribution of tablets and data capture was transferred to on-line systems.

D. Cycle 2.

New elements or 'Bonus Games', offering additional practice in transcoding and number-line completion, were introduced at regular intervals as participants proceed through the levels of the core game. These were intended to enhance motivation for the player and to explicitly reinforce key aspects of number learning, in particular the number sequence and the link between the spoken and written form of numbers.

A further important development to the game, enhancing access for all users, was the inclusion of animated tutorials preceding each level, giving an active demonstration of the specific demands to be made of the player (see below for an example). These aimed to increase players' independence in progressing through the game.

A screenshot from a video tutorial. The finger has moved through the actions required to initiate and follow a sequence from 25. It will go on to press 28 twice in order to complete the sequence.

Based on findings from Cycle 1, changes were implemented to increase sensitivity of assessment measures. For transcoding (reading, writing and identifying Arabic numerals), the range of items was extended. Assessment of counting aloud was modified to include counting on and down from specific target numbers. The Picture Analogy Test (PAT), a non-numerical control task, was developed, drawing on research findings and assessment materials commonly used the United States. Assessed before and after intervention, performance on this task should not be affected by experience with the SWAN game. Cycle 1 results showed ceiling effects on the PAT, so several more difficult items were added. Pilot testing indicated that the materials would be appropriate for both children with DLD and people with aphasia.

Revised assessments of Functional Numeracy were developed for PWA and children with DLD. For the PWA version, the range and difficulty level of the procedure were extended, and trials were conducted with neurotypical adults. The Functional Numeracy assessment for children with DLD was modified to provide a stronger link between visual and linguistic stimuli, and a more gradual progression of difficulty. The resulting material was revised following piloting with typically developing children.

The move to online delivery of the study was a major undertaking. Combined operation of Gorilla, an online experiment building platform (https://gorilla.sc/), with Zoom video platform (https://zoom.us/) was identified as the preferred approach. Graphic presentation of the existing number knowledge assessments for online delivery was relatively straightforward, given the uniform presentation of items. Work on graphics for functional numeracy assessments was more time consuming as it required item-specific stimulus presentation, especially in the children's version. Eventually it was decided that this assessment, alone, should operate as a paper-based task mailed to participants for completion pre- and post-intervention. For all the Gorilla-based tasks response capture and processing of raw data, which was specified to include response times, required complex and time-consuming programming.

The wider geographical scope offered by online delivery proved valuable, and recruitment for PWA was encouraging. However, given the pressures on education during the pandemic it was understandable that the response from schools was sparser. It was decided to focus on trials with PWA at the start of Cycle 2 in early 2021, and to relaunch recruitment initiatives for children with DLD during February and March, based on direct approaches to families. The response was promising, but many of those expressing interest were unable, for various reasons, to participate.

Final participation numbers (Cycle 2) were as follows:

Eighteen people with aphasia (age range 37-85 years) completed the study, including post-intervention and follow-up assessment. Six children with DLD (age range 5 years 9 months to 7 years 10 months) completed the study, including post-intervention assessment, but were not available for follow-up assessment due to the continued impact of the pandemic on schools and families.

Of the six children with DLD, one had an additional diagnosis of 7q11 duplication syndrome – her primary difficulty was with speech output. Four of them worked remotely from home, two worked remotely from school.

E. Results.

Assessment measures included: transcoding (number identification, number reading, number writing), counting aloud (forward and backward) and calculation (addition without carrying, addition with carrying, subtraction without carrying, subtraction with carrying), and a non-numerical Picture Analogy Test (PAT), a test of analogical reasoning. PAT was used as a control task which was unlikely to show effects of intervention.

All assessment measures were taken twice (baseline 1 followed a week later by baseline 2) before the SWAN game was given to participants. Following three weeks of gameplay (for around 15 minutes per day), assessments were carried out for a third time (post-intervention). For PWA only, assessments were carried out for a fourth time six weeks later (follow-up, post-intervention 2).

People with Aphasia

All baseline measures showed good reliability and were sensitive to the wide variation in numeracy skills showed by PWA. Performance levels in basic number processing (transcoding and counting) lent themselves to categorisation as follows: relatively good skills, 4 participants; poor skills, 7 participants; mixed profile (strengths in transcoding, poor counting), 3 participants; mixed profile (islands of strength and weakness), 5 participants.

These number processing categories proved to be strongly associated with the severity of aphasia of the participants. PWA whose number processing was poor had significantly more severe aphasia than the rest. More detailed analysis confirmed linkage between specific numerical and linguistic skills. Object naming skills were highly correlated with number reading, and auditory word processing was highly correlated with number identification (matching spoken numbers to Arabic numerals).

Intervention (playing the SWAN game) produced widespread positive effects on number processing (transcoding and counting) and calculation. Only one participant failed to make any progress attributable to intervention. Five participants showed intervention-based improvements on one numeracy task. Six participants improved on 4, 5, 6 or 7 tasks, including both number processing and calculation. Calculation appeared to be more vulnerable than number processing to showing immediate post-intervention gains which were not maintained at follow-up.

Gameplay characteristics varied considerably between individuals with PWA. Many, especially those with relatively high number processing skills, completed all 140 levels. Other individuals also completed all, or the great majority of levels. While this extensive exposure to the game was broadly related to outcomes (growth in skills), there were numerous exceptions. The participant with the highest number of good outcomes reached only level 40 of the game but did so at a slow rate. The benefits of playing the SWAN game were highlighted by analysis of the gameplay data, which revealed associations between the number sequences that received mass practice and the gains individuals made in number processing post intervention. For example, participants who worked extensively with three-digit numbers were able to name/identify these numbers more accurately.

The SWAN game was well-received by PWA. Many participants wished to continue using the game after the intervention period was completed.

Children with DLD

With one exception (the participant with 7q11.23 duplication), children with DLD showed far lower performance levels in numeracy tasks than the majority of PWA. They also experienced more problems in engagement with the online assessment process. In several cases attentional difficulties had been noted in clinical reports. In one case (AD20) the stress of assessment combined with extreme technical difficulties meant that reliability of measurement became uncertain.

Delays due to COVID meant that participants were only available for immediate post-intervention assessment. Thus, we were able to identify growth in skills attributable to intervention, but unable to test for maintenance of any such growth.

Only AD20, for whom assessment was unreliable, failed to show improvement on at least one measure of numeracy. One participant whose baseline measurements were extremely low, showed significant and substantial improvement in counting. Other participants showed improvement on two or three measures, largely limited to number processing tasks. Only one participant showed improvement in calculation (addition without carrying). It should be noted that the main assessments of calculation, while allowing direct comparison with performance levels of PWA, made demands which often exceeded the skills and experience of children with DLD.

Gameplay characteristics of children with DLD (except for FD24), differed substantially from those of PWA. Most only reached game levels in the 20s or below. Unlike PWA, they frequently failed to reach criterion for progression to the next level, often requiring numerous repetitions before advancing. However, meeting these challenges appears to have had positive effects. For example, teen numbers which caused difficulties at pre-intervention assessment were given repeated supportive practice in gameplay and were no longer problematic at post-intervention assessment. The persistence with which players with DLD maintained attention, even in cases where attentional issues had been previously reported, was striking. Informal feedback confirmed these findings for two children in particular. Both, once provided with the tablet and basic instruction, were able to concentrate on the game in future sessions, without additional management. Both made significant gains post-intervention. Both were sad to learn that the tablets had to be returned on completion of the project.

F. Conclusions

The study provided strong indications that the SWAN game may be used as an intervention to enhance basic numeracy in PWA. Similar indications were found for children with DLD, though the data were sparse. The study has confirmed the importance of unmet needs for numeracy support in PWA and provided a possible framework for categorization of such needs. In both PWA and children with DLD, improvements attributable to intervention were observed across the range of pre-intervention numeracy levels and gameplay skills. There are indications that the learning algorithms written into the SWAN game were successful in supporting participants with a wide range of learning difficulties.

Further data needs to be collected to establish whether these findings can be replicated or indeed surpassed when the intervention is being delivered in less challenging conditions. We still need to determine what is an ideal gaming experience – time, using headphones to limit distractions, social engagement etc.

It is recommended that future research focuses on developing the SWAN game so that it targets the needs of different client groups more effectively. Some children with DLD appear to need a slower progression and more support in the early stages of the game; while PWA, might benefit from a wider range of bonus games to support targeted computational skills and encourage progress.

Of great value would be the development of algorithms that would allow players to progress through the game according to their abilities. Individuals with more advanced skills could enter the game at a higher level and focus more on challenging number sequences.

Other adaptations to the game, including voice recognition, would allow us to investigate whether both output and input processing need to be targeted if individuals are to make maximal gains.

The SWAN project has generated interest amongst researchers in the field as well as colleagues in health and educational settings. Future work could also include trials with individuals with poor number skills who do not have language problems.

2. Introduction

2.1 Research Background

Numeracy is essential for educational success and participation in society. Recent research highlights the importance of foundational number language in building arithmetical skills (Habermann et al., 2020). The number sequence itself, in spoken and written forms (number words and Arabic numerals) provides the basis for the countless numerical activities each of us performs every day, in the playground, in the classroom, in the office, in the kitchen, and so on. As well as enabling calculation, spoken and written numbers provide a specialised and highly efficient system of communication. In childhood, shared counting with an adult or another child provides a perfect framework for learning through interaction (Durkin et al., 1986). As spoken numbers are practised and memorised, so, gradually, is their correspondence with the Arabic numeral system (Yuan et al., 2019). Conceptualized in this way, the acquisition of basic numeracy can be understood as a specialised form of language learning. It is unsurprising, therefore, that individuals with developmental and acquired language disorders often have problems with numbers.

Around 8% of children in the UK have a developmental language disorder (DLD). Several research studies have shown that children with DLD often have difficulty acquiring basic numeracy skills. For example, Donlan et al. (2007) found that 88% of a sample of eight-year-olds with language disorders performed below the level of their age-peers on a range of mathematical tasks. These children had particular problems in production of the spoken number sequence and in 'transcoding', i.e., producing the spoken forms of Arabic numerals, and identifying an Arabic numeral from its spoken form. Despite these difficulties, the study found that the children with DLD had relatively good underlying number concepts, suggesting that, if foundational number processing skills could be firmly established, enhanced mathematical development might be expected.

The relation between number processing and language processing in adults is a matter of intense debate. Sometimes it appears that number skills and language skills are independent. The film 'Rainman' famously depicts exceptionally advanced numerical abilities in an autistic man who has significant language and communication difficulties. Research has shown that some individuals with aphasia (impaired language following brain injury) have well preserved mathematical skills (Varley et al., 2005) even though they cannot process comparable rules in language and have problems understanding number words, such as ‘three’. This finding suggests that, at least for some adults, linguistic and mathematical processing do not make use of the same cognitive reserves. However, it appears that this is not a general pattern. Evidence is lacking concerning the prevalence of numeracy difficulties amongst people with aphasia (PWA), but a combination of published research and clinical experience suggest that a high proportion are likely to have substantial difficulty with number processing, and that this may be more prevalent than in the general population (De Luccia & Ortiz, 2016). One study found financial and consumer issues, such as accessing telephone banking services and buying a ticket, to be the most frequently raised concerns amongst PWA (Morris, Ferguson & Worrall, 2014). Despite these indications of the significance of numeracy difficulties for PWA, intervention to address these issues is rarely undertaken.

The commonality of effects on basic numeracy in children with DLD and adults with aphasia, as reported above, is striking. However, these are areas of research and practice which are pursued, for the most part, independently. Motivation for the Sequences in Words and Numbers (SWAN) project derived from interactions between team members whose practical and research experience with the two populations exposed the common ground. The SWAN project sought to examine the underlying constraints on numeracy by developing and evaluating an evidence-based intervention applicable to both populations.

2.2 SWAN Prototype

A very early prototype of the SWAN game had been proposed by a client with aphasia who attended the UCL Communication Clinic and had marked difficulties with number naming and sequencing. He wrote a basic computer programme, based on research evidence, which provided multiple opportunities to link symbols with their spoken names within a sequence. He demonstrated that entering the sequence at points beyond ‘one’ was effective in helping him consolidate his knowledge. However, it was a repetitive activity that required self-motivation to engage with the task and the programme was not visually attractive.

The SWAN team developed this early prototype in two important ways. First, the notion of a training programme based on simple repetition should be modified through ‘gamification’, i.e., the target behaviours (number sequence identification and association of spoken vs. Arabic forms) should operate as core activities within a computer game, rather than a decontextualised training programme. The techniques of the gaming industry, including sophisticated artwork and reward systems, should engage the user in a gaming activity which is intrinsically motivating, thereby facilitating repeated distributed practice of the target behaviours over several sessions spaced out over time. Countless studies have shown that verbal learning is greater following practice that is distributed across multiple sessions rather than completed in a single session (e.g., Middleton, Rawson & Verkuilen, 2019). Second, the game should entail a developmental progression. The number sequences on which it is based should increase in range and complexity in a way which corresponds to the pattern observed in children's learning (Fuson, 1988). Anecdotal evidence suggests that this pattern of counting is mirrored in people with aphasia (e.g., with some individuals unable to enter the number sequence at points beyond one). The game should serve as an intervention to support basic number knowledge in both children with DLD and people with aphasia.

The SWAN game prototype, SWAN Version 1 (V1), was developed according to these principles, with seedcorn funding from the Institute of Cognitive Neuroscience at UCL and UCL Business, in collaboration with the gaming software company SoftV. The prototype offered an attractive graphic interface and was programmed to follow a simple progression through 70 levels of mastery of number sequence identification. Every time the player touched a number on the screen the game would produce the corresponding spoken form. SWAN V1 also included a parallel form based on the alphabetic sequence. Following informal trials and information gathered from a focus group the prototype was modified to provide enhanced graphics (see Figures 1 and 2), and a more structured interface with the player. Fixed gameplay rules setting the criterion for access to successive levels were established through collaboration between the UCL team and SoftV programmers. These revisions were incorporated in SWAN V2. Pilot studies were then conducted with typically developing children (Long, 2018), and with PWA (Bruce et al., 2017). Although results indicated that the game might have utility as an intervention to support basic number knowledge, they also highlighted the fact that significant modifications to the game would be required to accommodate individual differences in number knowledge and rates of learning observed in PWA, and likely to be encountered in children with DLD. SWAN V2 would require that players should be required to 'pass' each level in order to progress. Furthermore, in order that gameplay itself be examined as a factor in subsequent analysis, a system of usernames was required to support individual gameplay data capture, and transmission of these data to a remote server.

Figure 1.

The first 'board' at Level 26 of the SWAN game. The player must touch adjacent tiles in consecutive order. Scores are based on the length of sequences entered.

Figure 2.

The SWAN 'world' for levels 11-20. This player must reach criterion score on level 11 in order to proceed to the next level.

The current feasibility study sought to implement these changes through collaboration with SoftV, and to produce SWAN V3, which would allow us to evaluate the possible utility of the SWAN game as an intervention to support basic number knowledge through a series of single case intervention studies carried out with samples of children with DLD and adults with aphasia.

3. Project Aims

- To develop a sequel to the SWAN prototype with a range of new features designed to improve capture and analysis of individual gameplay data (including machine learning approaches), and to improve ease of access for the player, tailored to the needs of (a) children with developmental language disorders and (b) adults with aphasia, and suitable for use in a future randomized control trial.
- To record the development of number sequence, transcoding and arithmetic skills, and their practical application, at an individual level in participants from both study groups.
- To test the feasibility of using adaptive programming, based on machine learning, to extend the flexibility of the SWAN game in providing a responsive learning environment across the range of knowledge and skills represented in the two user groups.
- To evaluate the effectiveness of the SWAN app in enhancing everyday numeracy using a single-case series approach.

4. Study Design

4.1 Joint Working

An overarching element of the study design concerned collaborative working with SoftV. A positive relationship had already been established during the production of the prototype (SWAN V1) and modifications thereof (SWAN V2). Over the course of the study, the intention (before the advent of COVID-19) was that two cycles of activity would be conducted, each of which included phases of game development, intervention (single case intervention studies), and evaluation. In each cycle, ten single case intervention studies would be undertaken with each participant group (children with DLD and PWA) (see Table 1). Once Cycle 2 was completed, a final development stage would address remaining modifications required for the game to be utilised in a large-scale study.

Table 1: Study Cycles and Sample Sizes.

Cycle 1	Game Development	Intervention	Evaluation
	Prepare Assessments	n=10 DLD, n=10 PWA	
Cycle 2	Game Development	Intervention	Evaluation
	Modify Assessments	n=10 DLD, n=10 PWA	

Joint working on game development and evaluation would be central to the success of the project. In the first stage of Cycle 1 the UCL SWAN Team and SoftV would implement changes to the game as indicated by pilot studies, thereby producing SWAN V2. Following the intervention stage, gameplay outcomes would be jointly evaluated, providing indications for further development leading to SWAN V3, to be presented to participants in Cycle 2.

4.2 Intervention Design

Assessments of basic numeracy, based on Donlan et al. (2007) and Habermann et al. (2020), were compiled during the Game Development phase of Cycle 1, and closely matched for children with DLD and PWA, to be administered twice (baselines 1 and 2; b1 and b2), with an interval of one week, with the intention of establishing a stable baseline against which any intervention-based change might be measured. Assessments would be repeated immediately post-intervention, and again after six weeks (post-intervention 1 and post-intervention 2; p1 and p2). Our study distinguishes between assessments of proximal goals, i.e. skills which closely match the behaviours entailed in playing the SWAN game, and distal goals, those which are more distantly related to the behaviours entailed in the gameplay. Assessments of proximal goals here are grouped together under the general term number processing, and comprise tests of transcoding (number identification, number reading and number writing) and tests of counting (forwards and backwards). These tests of basic number processing tap the foundations of functional numeracy and underpin more complex procedures. Distal goals within the study are grouped under the general term calculation and comprise addition (with and without carrying) and subtraction (with and without carrying). The assessment battery also included a non-numerical control task, based on analogical reasoning and testing the ability to detect meaningful patterns. The Picture Analogy Test (PAT) was developed for use by both groups. This task was based on research findings and assessment materials commonly used the United States. Performance on this task was expected to remain stable, thereby providing an indication of the specificity on any post-intervention change in numerical skills. New assessments of contextualized functional numeracy developed for the study, with different versions for children with DLD and PWA. These were administered once before intervention and once after intervention.

Following baseline assessments, the SWAN game would be introduced to the participants. They would then play the game, at home (PWA) or at school (children with DLD), for 10-15 minutes per day for three weeks. During this time, gameplay data would automatically be sent to a remote server where it would be available for purposes of compliance monitoring and data analysis. Assessments would be repeated immediately after the intervention (Post 1), and again after a six-week interval (Post 2), to evaluate maintenance of any post intervention changes.

5. Cycle 1

The initial phase of Cycle 1, including game development and preparation of assessments, took place between October 2019 and February 2020.

5.1 Game Development

5.1.1 Enhanced Progression

Substantial changes were made to the game to increase its accessibility for participants with limited basic number knowledge, and to extend the coverage and difficulty levels for those PWA for whom further challenges would be required. Game re-design (SWAN V3) addressed both concerns. The initial levels of the game were simplified in terms of board size, number range and progression. The 70 levels of the prototype were expanded to 140. Table 2 gives an indication of which game parameters are manipulated within the progression through levels.

Table 2. Examples of variation in board size, sequence type and range according to levels of SWAN V3.

Level	Board Size	Sequence Type	Number Range
1	3x3	Ascending	1-3
20	6x6	Ascending	13-20
50	5x5	Descending	7-1
100	7x7	Ascending Even Numbers	2-52
140	4x4	Ascending in 5s	5-20

The final board from Level 1 of the revised version of the game is shown in Figure 3. In Figure 3, the player has pressed tiles in the sequence of 1-3 and heard each of the corresponding spoken numerals produced by the tablet. Now the player should press the final tile again to indicate a complete sequence and complete their action. The entered sequence will then disappear, adding points to the score according to the length of the sequence. This player has already accrued 27 points from previous boards. Once the current sequence is cleared, it will disappear from the board, leaving tiles which allow two further sequences of adjacent tiles to be entered. These basic principles of gameplay had been established in V2. Note, however, that this third board at Level 1, based on the simple sequence 1-3, exploits the principle whereby adjacency is defined vertically, horizontally and diagonally, requiring the player to identify sequences which do not conform to canonical left to right ordering. This exemplifies the more gradual progression in gameplay introduced in V3.

Also notable in Figure 1 are the 'help' arrows which appear when the player is slow to start their turn. The help system was revised and expanded in V3, to support players with limited basic number knowledge without hindering the progress of more skilled players.

Figure 3.

A board from Level 1, showing score from previous boards and 'help' arrows.

At Level 2 each board has a 4x4 matrix, including possible sequences in range 1-4. The progression over the remaining 138 levels includes board sizes up to 7x7 (see Figure 4) and covers ascending and descending sequences, including sequences of 10s, 5s, even numbers, odd numbers, 3s (see Figure 5), and 4s. A demonstration of various levels of the game can be found at https://www.youtube.com/watch?v=1hy4k3IcaOQ

Number ranges within 1-99 are included, with subsets focusing on teen numbers and decade boundaries, based on findings from developmental research. The final progression represents both the accumulation of number sequence knowledge but also the expansion of gameplay skills. Board size increase is linear but varies according to number range and type of sequence. V3 also incorporated constraints on sequence construction, introduced between Levels 4 and 8, including, filler tiles ('water-lilies') and blank tiles which can stand for any number in a sequence. 'Shell' tiles bear a number but are distinguished from standard number tiles since they can be re-used, and don't disappear once the sequence within which they are used is cleared. Re-use of the shell increases the player's score. Deliberate use of shells indicates a higher level of skill.

Figure 4.

A board from Level 89, showing a 7x7 matrix and including blank tiles (lily pads) and a shell, which may be used in more than one sequence, thereby increasing the player's score.

Figure 5.

The first board from Level 107, a 5x5 matrix introducing counting in threes.

5.1.2 Joint working

The revised version of the game entering intervention trials in Cycle 1 represents successful collaborative working between researchers and gaming engineers. Although the game concept is elegantly simple - the player produces number sequences of increasing complexity, based on developmental research - the implementation of this concept is far from simple. The expertise of gaming engineers is needed to generate a user interface which maintains surface simplicity and provides, at the minimum, easy access, an engaging graphic environment and an effective reward system. Skillful combination of these factors should provide intrinsic motivation far beyond what could be achieved without ‘gamification’. The technical demands involved in implementing this progression in the game were significant. The non-technical specification of levels provided by the UCL Team for implementation by SoftV was itself complex, demonstrating the substantial sharing of ideas which had taken place between the teams.

5.2 Data capture

In terms of data capture, every press of a tile, its location on the screen and timing in relation to other actions could be captured as raw data, but these myriad datapoints would need to be carefully processed to serve the researchers’ aim of identifying individual differences in gameplay. From the research point of view, particular importance was attached to variation in the length of sequence entered by players, considered in relation to the longest possible sequence available on the screen. A count of types of errors made, i.e., tiles pressed which did not satisfy (a) the requirements of adjacency or (b) consecutive order, was also considered important, as was a record of engagement of the help system. Each of these measures was to be taken for each sequence entered and summarized for each level.

However, discussions between the UCL and SoftV teams concerning data capture demonstrated the challenges of collaborative working. While sequence length reporting was successfully negotiated, the recording of error and use of help posed significant difficulties. Defining and distinguishing the raw data entries which would register error types and use of help and would calculate time spent engaged in gameplay at each level, proved to be much more difficult than anticipated. A possible solution was provided by SoftV in the form of a Python script (computer code) which was to be used by the research team to aggregate gameplay data downloaded from the server, and intended to present summaries of error type, use of help and engagement time. Communication proved to be difficult. Uncertainty concerning the output of the script persisted. A further issue concerned the attempt to use machine learning techniques to identify patterns of user performance. Again, communication proved to be problematic, and it was decided that energy would be best devoted to the simpler task of achieving accurate data summaries.

5.3 Effects of COVID -19

The SWAN project was affected by the COVID-19 crisis in a significant way. Primary data gathering for Cycle 1, was planned to take place through March, April and May of 2020, and to include 10 adult participants with aphasia, and 10 children with DLD. However, this was curtailed by lockdown on 23rd March which closed schools and prevented PWA from travelling to assessment appointments. The status of Cycle 1 at lockdown was as follows:

PWA (Target 10 participants):

- Complete baseline measurements on six participants
- Partial baseline measurements for two participants
- Complete gameplay data on six participants
- Post-intervention measurements (but no follow-up measurements) on six participants

Children with DLD (Target 10 participants):

- Complete baseline measurements on seven participants
- Less than 30% gameplay data on seven participants
- No outcome measurements

Overall, targets for participant numbers were not reached, gameplay data capture was largely truncated, and we were unable to carry out the majority of post intervention assessments.

While it bears no comparison with the devastating effects of COVID-19 on individuals, families and society as a whole, the effects of the pandemic on the SWAN project were profound and long-lasting. It soon became clear that Cycle 1 could not be completed, and that face-to-face working, on which all assessment procedures were based, was unlikely to be possible in the foreseeable future. Coronavirus restrictions continued to operate in schools and health centres, preventing face-to-face contact with participants. SWAN staff took the decision to pause research activity for two months, in the hope that effects of the pandemic would diminish during that period. After two months, existing Cycle 1 data would be analysed in order to guide preparations for an extended Cycle 2. An expanded sample (n=15 per group, instead of n=10) would be sought, and the entire procedure of assessment, distribution of tablets and data capture would be transferred to online systems.

5.3.1 Data Recovered from Cycle 1

While the data from Cycle 1 were far from complete, especially regarding outcome measurement, baseline measurements were taken from 8 PWA and 7 children with DLD, all of whom met our criteria for inclusion. Analysis of these data identified the pattern of error in transcoding (reading, writing and identifying Arabic numerals) expected in children with DLD, while PWA (whose numeracy skills are largely unresearched), showed a contrastive pattern, some responding correctly but needing time to get to the target, some producing errors and self-correcting. Individual differences within both groups were substantial, including variation across tasks within participants as well as differences between individuals. These indicative findings provided support for the contrastive element of the study design in which initial learning challenges (for children with DLD) are compared to re-learning challenges (PWA). Item analyses conducted across all baseline assessments identified improvements to be made in advance of Cycle 2. For transcoding, these included expansion of the range of items to avoid ceiling effects, and, if technically possible, response time capture (particularly for PWA). For counting aloud, changes were needed to enhance sensitivity and reliability of scoring. The non-verbal control task showed ceiling effects. Functional numeracy tests (which differed for children with DLD vs. PWA), newly developed for the project, presented challenges both in terms of difficulty levels and in the linguistic demands of administration, and required comprehensive review.

Six PWA completed gaming as planned. The seven children with DLD had completed only a small proportion of planned gameplay when lockdown prevented further engagement. Contact with families of two children with DLD was made to allow home use of the game for a period over the summer vacation. Approaches were made to AFASIC (https://afasic.org.uk) to capture further gameplay data from children with DLD attending summer schools, though, despite best efforts on all sides, this was not successful. Where contacts had been made, remote data capture, utilizing home wi-fi for PWA and school wi-fi for DLD was successful for both groups. The technology overall appeared to be reliable and trouble-free for users. Primary data were successfully recovered from the server. More problematic was the secondary data (error rate, use of help, time per level) generated by the Python script (see 5.2, above). There were doubts about accuracy which were referred to SoftV. An analysis of primary data, comparing gameplay of two people with aphasia (representative of the range of performance in the group) against the gameplay of two children with DLD showed significant variation in rates of progress through the game (see Table 3).

Table 3. Individual differences in rates of progress through the game.

Participant	Final Level Completed (Max. 140)	Board Size	Level Content	Number of Sequences Entered	Number of Days Played
PWA 1	100	7x7	Ascending 2-26 in twos	1253	17
PWA 3	74	5 x 5	Descending 53-45	725	18
DLD 1	9	6 x 6	Ascending 1-10	103	4
DLD 6	21	4 x 4	Ascending 10-50 in tens	322	5

The game was extremely well-received by both groups. Participants with aphasia all commented that they enjoyed playing. They usually played for twice the amount of time suggested, often for 30 minutes or longer each day. Teachers of children with DLD commented '…the kids loved the tablets and game…', '…He is so motivated and loves doing it every morning….'. The fact that motivation was high, and maintained across multiple play sessions, was an important finding. This is particularly important since the content of SWAN is entirely focussed on the number sequence and the association between Arabic and spoken forms and is designed with the principle of mass practice (which has proved effective in other intervention studies) at its heart. Thus, any engagement with the game provides input relevant to participants' needs.

6. Cycle 2

Cycle 2 began in September 2020, even though at that time schools and health centres remained closed to non-essential staff, and teachers and speech and language therapists working with children with DLD continued to face considerable additional pressures. These pressures continued throughout the length of the project.

6.1 Game Development

Given the positive feedback received from users in Cycle 1, no further modifications to the interface features or the progression within the game were indicated. Instead, new elements or ‘Bonus Games’, offering additional practice in transcoding and number-line completion, were introduced at regular intervals as participants proceed through the levels of the core game (see Figures 6 and 7). These had been in development before Cycle 1 but had not reached completion. Research places special emphasis on transcoding (Malone, Burgoyne & Hulme 2020) and number line knowledge (Siegler 2016) as foundational elements of numeracy. The bonus games were intended to explicitly reinforce these key aspects of number knowledge, as well as to enhance motivation for the player. The bonus games would also provide an opportunity to monitor growth in number knowledge. The post-it game was inserted within the progression of the main game at levels 20, 40, 50, 70, 90, 110 and 120, using numbers corresponding to those in use in the main game at that level. The number-line game was inserted at levels 10, 30, 60, 80, 100, 130, using number sequences corresponding to those in use in the main game at that level.

Figure 6. The Post-It Game.

As each letter is presented for posting, a spoken number is produced (e.g. "seventeen"). The player then posts the letter using the touch screen. The number choices are designed to represent visual and auditory similarities which pose challenges in development.

A demonstration of the Post-It game is available here: [*https://www.youtube.com/watch?v=WSlfQj*](https://www.youtube.com/watch?v=WSlfQj)

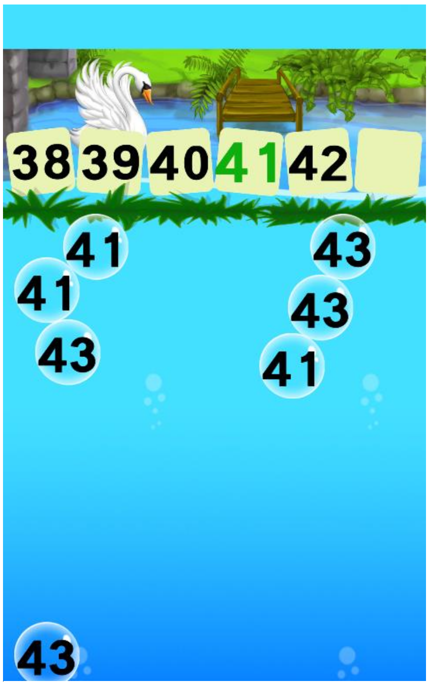


Figure 7. The Number Line Game.

Bubbles containing numbers rise from the seabed. Players must 'grab' the number needed to complete the sequence represented above and place it in the appropriate slot. Sequence ranges match the main game levels. Difficulty increases as the number of blanks increases and the interval between bubbles appearing decreases.

A demonstration of the Number-Line game is available here:

https://www.youtube.com/watch?v=-o2qCPcJGv4

A further important development to the game was the inclusion of animated tutorials preceding each level, giving an active account of the specific demands to be made of the player (see Figure 8 for an example). These were intended to enhance players' independence in progressing through the game. As with the bonus games, tutorial support of this sort had been planned at an earlier stage but was not developed in time for use in Cycle 1.

Figure 8. A screenshot from an animated tutorial. The finger has moved through the actions required to initiate and follow a sequence from 25. It will go on to press 28,29, 30, 31, 32 and then press 33 twice in order to complete the sequence.

The incorporation of new elements of content within SWAN V3 entailed rewriting the computer code, i.e., the programme on which the game is based. Rewritten code in one part of a programme can produce unforeseen errors, or 'bugs' in other parts. This proved to be the case with SWAN V3, where bugs were found in the operation of the main game. 'Debugging' i.e., detection and rectification of errors required intensive collaborative work. Further work was required to address data capture and processing issues, where communication between the research team and the data analyst proved to be difficult, as noted in 5.2, above.

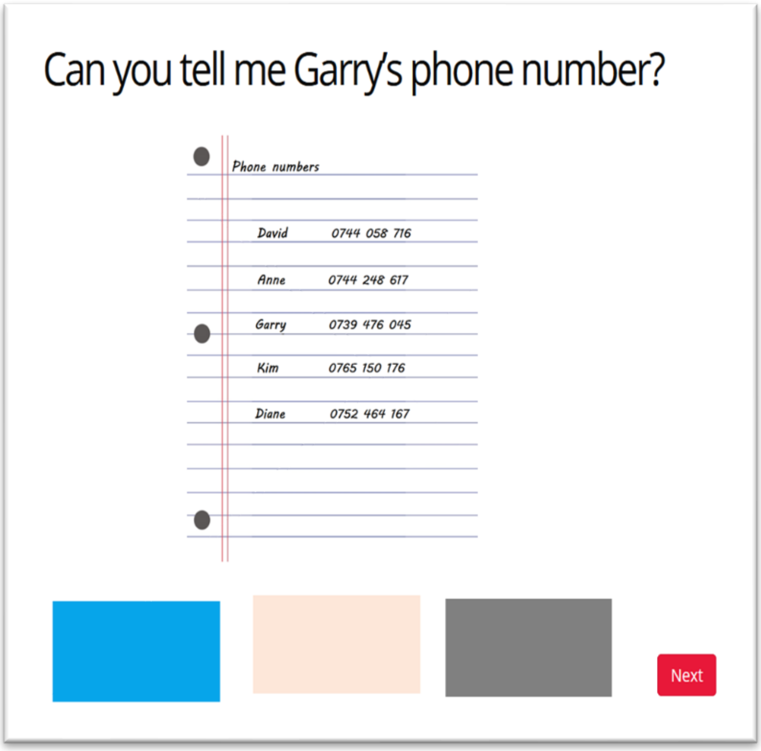
6.2 Assessments

6.2.1 Modification of assessment content.

As noted in 5.3.1., changes were implemented to increase sensitivity of measurement. For transcoding (reading, writing and identifying Arabic numerals), the range of items was extended to include further variation in double digit trials and additional triple digit trials, thereby addressing the problem of ceiling effects. Assessment of counting aloud, based on Korpipaa et al. (2020) was modified such that recitation of the complete sequence 1-41 was replaced with counting on and down from specific target numbers, thereby testing participants' ability to cross critical boundaries (decades and hundreds), and offering the opportunity to enhance reliability and sensitivity by using a scoring system based on single items.

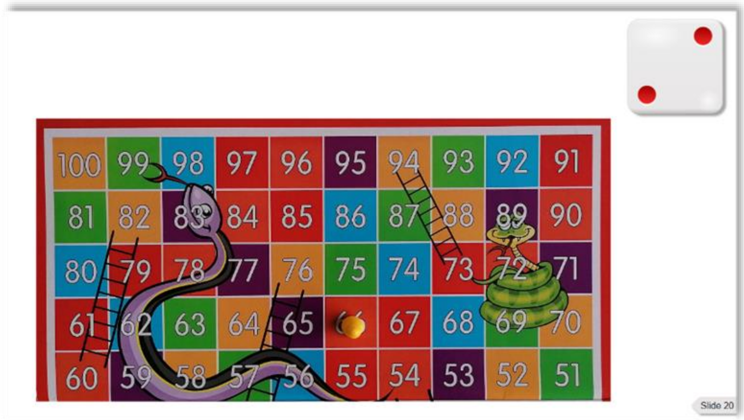
Revisions were made to the assessments of Functional Numeracy used in Cycle 1. For the PWA version, a group of high performing undergraduate and postgraduate students from UCL's Department of Language and Cognition extended the range and difficulty level of the procedure and trialled it with neurotypical adults. Based on a novel range of practical everyday tasks and including a range of formats of number representation encountered in everyday life, the assessment offers the opportunity to examine in detail the impact of loss of numeracy skills on everyday life. An example test item is shown in Figure 9.

Figure 9. Example item from functional numeracy test (PWA).



The Functional Numeracy assessment for children with DLD was revised by the SWAN team in line with findings of an item analysis conducted on data from Cycle 1. The pattern of responses indicated a need for a better match between visual and linguistic stimuli, and a more gradual progression of difficulty. The resulting material was revised following piloting with typically developing children. Figure 10 shows the graphic for item 15 ("You throw the dice. What number have you thrown? Where will you land?").

Figure 10. Example item from functional numeracy test (DLD).



Revision of the PAT (non-verbal control task) was made to avoid ceiling effects. Additional items of a higher level of difficulty were added following piloting with typically developing children.

6.2.3 Development of online assessment capacity

Combined operation of the Gorilla, an online experiment building platform (https://gorilla.sc/) with Zoom video platform (https://zoom.us/) was identified as the preferred approach to collecting assessment data remotely. Graphic presentation of the number knowledge assessments for online delivery was relatively straightforward, given the uniform presentation of items. However, work on graphics for functional numeracy assessment was more time consuming as it required item-specific stimulus presentation and response capture, especially in the children’s version. Eventually it was decided that the children’s version of this assessment, alone, should operate as a paper-based task mailed to participants for completion pre- and post-intervention. For all the Gorilla-based tasks however, response capture and processing of raw data, including response times, required complex and time-consuming programming. The additional complexity required for Zoom-based presentation of the Gorilla materials presented further challenges, extending the time required for development.

6.3 Recruitment

Given the constraints of the pandemic, it was anticipated that recruitment of the revised target of 15 PWA and 15 children with DLD would prove challenging. Major recruitment initiatives were based around DLD Awareness Day (16th October 2020) and World Stroke Day (29th October 2020). Video material was prepared and broadcast in successive waves on social media, with strong support from members of the advisory group, and other team contacts. Widespread positive responses were recorded. The SWAN team participated at UCL's World Stroke Day forum, providing video material in advance, and running Q&A and workshop sessions for forum participants. Further recruitment material was circulated through individual contacts at Dyscover, AFASIC, and through the NAPLIC newsletter. Based on findings from Cycle 1 and taking account of the new demands made by online assessment, screening procedures were developed to check that potential participants conformed to criteria, and that facilities for online assessment were in place. In the few cases where participants had the necessary facilities, arrangements were made to download the game to the participant's own device; otherwise, a fully prepared tablet was delivered. The wider geographical scope offered by online operation proved valuable, and recruitment for PWA was encouraging. Give the pressures on education during the pandemic it was understandable that the response from schools was sparse. It was therefore decided to focus on trials with PWA at the start of Cycle 2 in January 2021, and to relaunch recruitment initiatives for children with DLD during February and March, based on direct approaches to families through the AFASIC website (The Association for All Speech and Language Impaired Children: https://afasic.org.uk/ and the E-DLD (Engaging with DLD: https://www.engage-dld.com) scheme. The response was promising, but as reported below, many of those expressing interest were unable, for various reasons, to participate.

6.4 People with Aphasia: Intervention

6.4.1 Participants

Nineteen individuals with aphasia were recruited to Cycle 2. All contacts and assessments were carried out online. Participant identification code, gender and age for all participants are presented in Table 4. One participant (A21) withdrew from the study after baseline assessment. All other participants completed the Western Aphasia Battery - Revised (WAB-R; Kertesz, 2007) to gather information about their language ability. This test is usually presented in person, so adjustments were made to enable assessment to take place online. Responses, based on subtests tapping spontaneous speech, comprehension, repetition and naming, were scored to yield an Aphasia Quotient (AQ). The AQ is a weighted average of all subtest scores relating to spoken language, measuring language ability, and classified according to categories of severity from mild to very severe. Both these measures are included in Table 4. Of the 18 participants who completed the programme, 13 were male and 5 were female. The mean age of this group was 62.5 years, the mean AQ was 73.1. Eleven were classified as mildly aphasic, three as moderate, three as severe, one as very severe.

Table 4. Participants: ID, Age, Gender and Aphasia Status

Participant ID	Gender	Age	Aphasia Quotient	Classification	Aphasia type
A20	female	47	92.9	mild	Anomic
A21	female	52	withdrew from study		
A22	male	70	24.3	very severe	Broca's
A23	female	48	96.1	mild	Anomic
A24	male	63	78.3	mild	Anomic
A25	male	85	84.4	mild	Conduction
A27	female	77	71.1	moderate	Conduction
A29	female	71	35.9	severe	Broca's
A30	male	72	92.6	mild	Anomic
A31	male	76	69.7	moderate	Conduction
A32	male	37	78.8	mild	Anomic
A33	male	61	84.4	mild	Anomic
A34	male	71	78.7	mild	Anomic
A35	male	64	45	severe	Broca's
A37	male	53	96.4	mild	Anomic

A38	male	71	61.5	moderate	Conduction
A39	male	39	93.2	mild	Anomic
A40	male	57	46.9	severe	Broca's
A41	female	63	85.5	mild	Anomic

6.4.2 Pre-Intervention Status

Variation between individual participants' scores is high. One participant reaches ceiling scores on all transcoding tasks, but not on counting or calculation. Others show floor effects on one or more tasks. A general tendency is found for scores to increase across baselines, although there is substantial variation, with numerous individuals showing a decline in scores from baseline 1 to baseline 2.

Correlational analyses show highly significant correlations between all number processing tasks (.75 to .88), moderate to high significant correlations (.50 to .76) between all calculation tasks and moderate to high significant correlations (.52 to .78) between all calculation and all number processing tasks, excepting only number writing, which does not correlate significantly with any measure of calculation. The newly developed functional numeracy test shows evidence of construct validity based on moderate to high significant correlations (.533 to .705) with all number processing tasks and all calculation tasks except addition without carry. PAT generally shows the intended dissociation from other baseline assessments. Only one number processing measure, number reading, is significantly correlated with PAT. None of the calculation tasks is correlated with PAT. A high significant correlation between PAT and functional numeracy is consistent with expectation based on the non-numerical contextualizing content of the functional tasks. In a large-scale school-based intervention study, Brookman-Byrne et al. (2019) reported that higher verbal analogical reasoning was associated with higher accuracy and faster reaction times in science and maths, and observed that relational reasoning skills may provide a common cognitive resource, whether applied in verbal-semantic or visuo-spatial contexts.

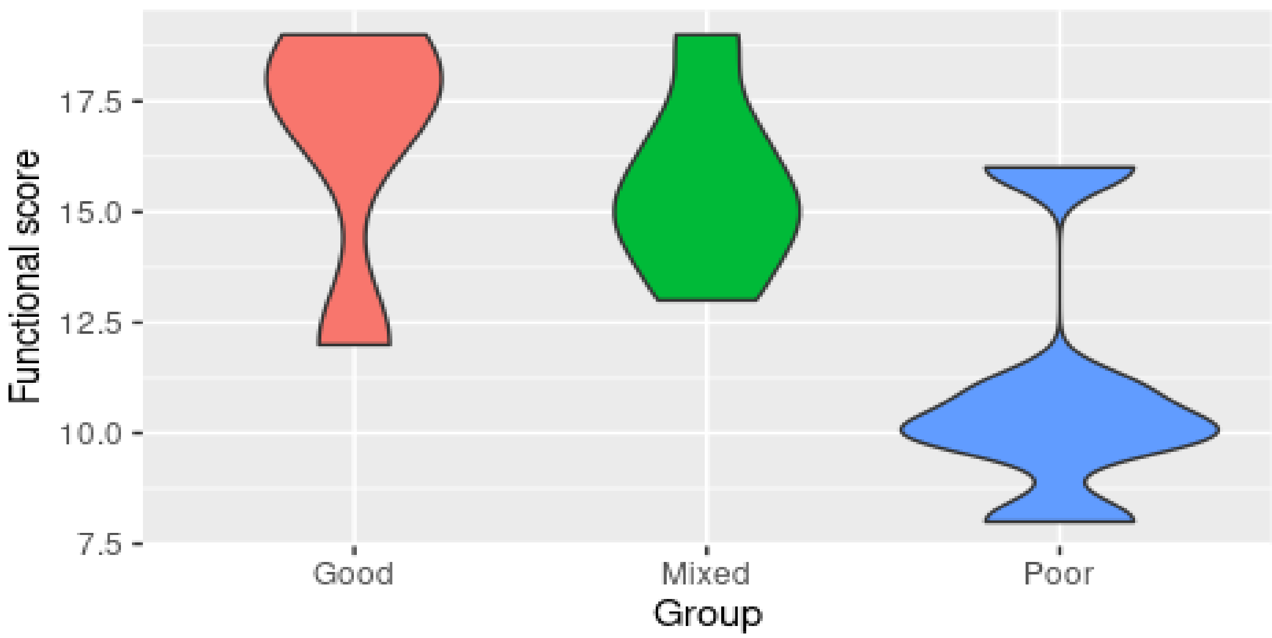
Notable is the finding that AQ is significantly correlated with all other measures, despite variation in levels of verbal loading, showing high correlations (.720 to .875) with all number processing tasks and with functional numeracy, moderate to high correlations (.528 to .778) with all calculation tasks, as well as a moderate correlation (.656) with PAT.

6.4.3 Number processing subgroups

Performance of PWA before intervention shows wide variation in number processing skill levels. Subdividing number processing into transcoding skills (number identification, number reading, number writing) and counting skills (forwards, backwards), clear subgroups emerge. Good (generally high performance) and poor groups (generally low performance) are readily identifiable. A further group has strengths in transcoding, but relatively weak counting skills. The final group shows a more complex set of profiles with islands of relative strength and difficulty. The two mixed groups are combined and shown as a single group in Figures 11 and 12.

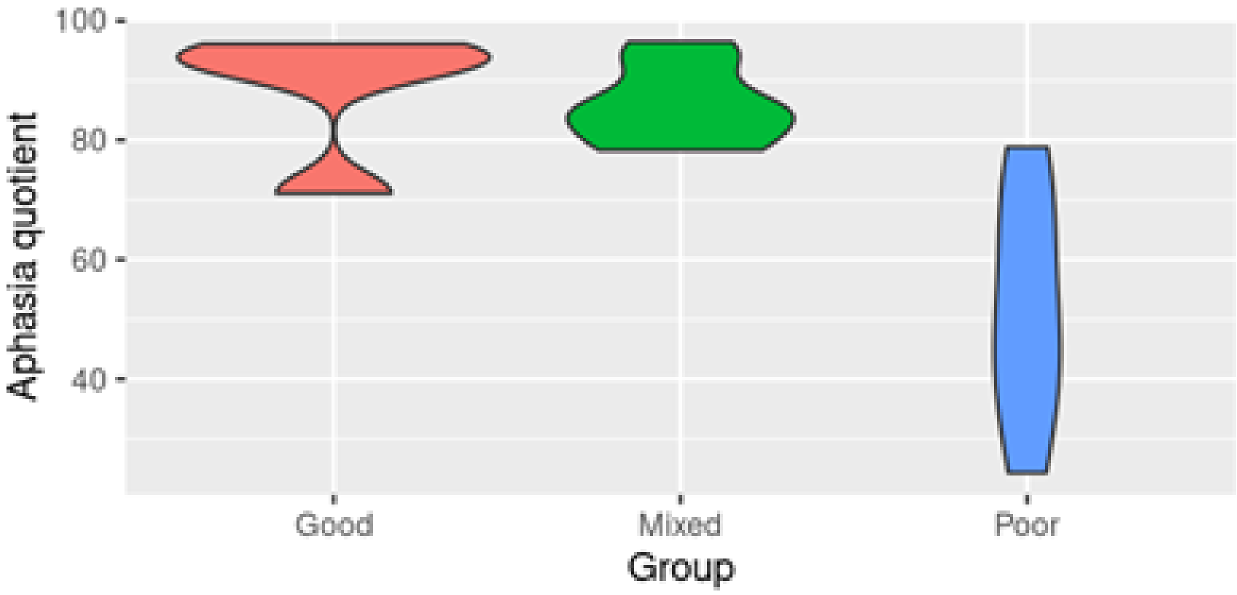
Evidence of construct validity in the functional numeracy test is provided by the association between number processing groups and functional numeracy scores (Figure 11). The great majority of the poor group scores fall below even the lowest scores of the good and mixed groups.

Figure 11. Violin plots of the relation between number processing groups and functional numeracy.



A similar pattern is observed in the association between number processing groups and aphasia quotients (Figure 12).

Figure 12. Violin plots showing the relation between number processing groups and aphasia quotients.



Exploring the association between aphasia and number processing in more detail, a significant high correlation was found between number reading and the WAB object naming test (both spoken output tasks). A further significant high correlation was found between number identification and the auditory word recognition subtest of WAB (both auditory input processing tasks). Both these results demonstrate a close correspondence between individual differences in specific aspects of the language of PWA and corresponding aspects of number processing.

6.4.4 Outcomes

In order to detect growth in proximal (number processing) and/or distal (calculation) skills attributable to intervention, data were analysed as a single-case series. The item-based weighted statistics method (WEST), designed for use in single case intervention studies, was employed (Howard, Best & Nickels, 2015). This approach takes account of differential responding across baseline measures.

Table 5 shows significant gains in assessment scores by number processing group. For A23 the opportunity to detect significant change in number processing tasks was removed by ceiling effects. Four participants (A27, A29, A38 and A40) showed gains in PAT as well as other measures, indicating a general improvement in performance which could indicate enhanced attentional skills, or to improved relational reasoning skills, or to other factors which may or may not be attributable to the intervention. The remaining 12 participants showed gains on number processing or calculation tasks, which may be attributable to intervention.

Most striking is the extent to which improvements are evenly distributed across groups. This gives a clear indication that the benefits of intervention are not limited by initial number processing status. Furthermore, within all groups, there was wide variation in the number and range of tasks showing improvement. Five participants improved on only one numeracy task. Five improved on 4, 5 or 6 numeracy tasks. These gains were not restricted to proximal (number processing) tasks. Nine participants improved in both distal and proximal tasks. However, it is notable that some individuals (A20, A32, A35) improved only in number processing while others (A22, A39, A40) improved only in calculation. The data suggest that calculation tasks may be more prone than number processing tasks to show immediate gains which are not maintained at follow-up.

Table 5. Results of WEST analyses showing gains in assessment scores, by number processing group.

Group/ Participant	PAT	Count fwd	Count bwd	No. ID	No. read	No. write	Add no carry	Add with carry	Sub no carry	Sub with carry
Good										
A23										
A27	✓			✓	✓	✓	✓	✓		
A30						✓			✓	
A39							✓			
Poor										
A22								✓		
A29	✓			✓						
A31		✓				✓		✓		
A32			✓							
A35					✓					
A38	✓	✓			✓	✓	✓		✓	
A40	✓			✓			✓			
Mixed profile -strengths in transcoding, poor counting										
A34		✓			✓		✓	✓	✓	
A37		✓	✓			✓		✓		
A41		✓	✓		✓		✓	✓		
Mixed profile - more complex										
A20		✓								

A21	NA									
A24								✓	✓	✓
A25		✓								
A33										

Note: ✓= Significant gain..

Empty cells: No evidence of gains.

Table 6 shows the range of effect sizes observed for each task. We bear in mind that our effect size measure is directly based on the number of specific items which show gains following gameplay. As such, the comparability of these measures is constrained by the comparability of the assessments themselves. Here there is high comparability within counting tasks, within transcoding tasks and within calculation tasks.

Table 6. Effect size range by task.

PAT	**Count fwd**	**Count bwd**	**No. ID**	**No. read**	**No. write**	**Add no carry**	**Add with carry**	**Sub no carry**	**Sub with carry**
-1 /+2	-5/25	-7.5/12	-2.5/+3.5	-2/+3.5	-4/+6.5	-1.5/+6	-4/+6	-4.5/+3.5	-4/+3

Notable are the wide ranges from negative values (items lost) to positive values (items gained) reported for all numeracy tasks. This finding reinforces the importance of single case methodologies in the evaluation of outcomes. Also of interest is the narrow range of values (around zero) observed for PAT. The pictorial presentation of this task, as well as its visual semantic content may explain its stability across pre-intervention and post-intervention measurements.

In the next section we will examine gameplay data and explore ways in which individual patterns of play might help to explain variation in intervention-based improvement.

6.4.5. Gameplay

Full gameplay data were captured for all participants who completed the study. One key element of the data, a record of the time spent by each player on each level of the game, was found to be inaccurate despite attempts to work around the problem. The following reliable gameplay variables were collected.

- Maximum level reached. The number of the final level played (out of 140). This is an indicator of the extent of exposure to the game, and to the skill level of the player. Since a criterion score (70% of maximum possible score) is required for progression, this variable represents a composite based on the number, length and complexity of number sequences identified by the player within the constraints of gameplay rules, and within the three-week intervention period.
- Number of repeated levels. The number of times a player failed to meet the criterion (70% of maximum possible score) for progression to the next level. The number of repeated levels is a general indicator of extent to which the game challenges the player’s gaming skill or number knowledge. Successful progression following repetition is an indicator of learning.
- Mean score per level. The average percentage of the maximum possible score achieved by the player at each level. This is an indicator of the player’s ability to manage the specific task demands (gameplay skills and sequence knowledge) of the levels they played.
- Mean time per tile. The average time taken (in seconds) for the player to select each tile in each sequence at each level played (excluding the time taken to select the first tile in each sequence). This is an indicator of rate of play and proficiency with the game.
- Mean errors per level. A measure of accuracy of gameplay. Errors are defined as failures to observe gameplay rules. These may represent gameplay error *per se* (e.g. selection a non-adjacent tile), or else number sequence failure (selection of non-consecutive number). Errors may also be recorded as a result of technical failure.
- Mean sequence proportion. This measure is based on comparison of the length of each sequence entered with the maximum possible sequence length available on the

current board. It provides an indicator of the player's understanding and engagement with the aim of producing the longest possible sequence.

Several individuals completed all 140 levels, but variation was substantial, with levels 50 to 130 represented with broadly uniform frequency. Repeated levels were rare. Only four participants repeated more than one level, though one of these repeated 11 levels. Most were easily able to reach criterion and progress (mean scores per level clustered above 90%). Consistent with these findings, time per tile (rate of play) falls generally between 1 and 2.5 seconds, but one outlier takes 4.4 seconds per tile. More extreme is the distribution of mean errors per level which is between 4 and 7 per level for the great majority, while one outlier reaches 20.26. Sequence proportion, on the other hand, shows uniformly high values, lowest of which is 90%. The overall picture is one in which, despite some ceiling effects, exposure to the game shows a wide distribution across the sample. Levels of skill and knowledge are concentrated towards maximum levels (or minimum, in the case of error and time per tile). However, there are exceptional individuals who fall outside this categorisation and are more challenged by progression requirements, less skilled (as indicated by rate of play), and more error prone.

Table 7 shows individual gameplay characteristics, and the number of numeracy assessments showing significant growth, as a function of number processing group.

Table 7. Individual gameplay parameters and outcomes by number processing group.

		Max. level	#Repeated levels	Mean score	Time per tile		#Improved outcomes[1]
Good							
	A23	140	0	0.95	1.22		0
	A27	140	0	0.93	1.92		6
	A30	140	1	0.94	1.66		3
	A39	140	0	0.95	0.78		1
Poor							
	A22	55	1	0.89	2.7		1
	A29	74	7	0.83	4.44		0
	A31	129	1	0.87	2.41		3
	A32	71	0	0.96	2.97		2
	A35	105	1	0.91	2.41		1
	A38	40	0	0.91	2.29		7
	A40	71	0	0.94	2.87		1
Mixed profile -strengths in transcoding, poor counting							
	A34	140	0	0.94	1.45		5
	A37	70	4	0.88	1.02		4
	A41	110	4	0.87	1.09		2
Mixed profile - more complex							
	A20	70	1	0.88	1.55		1
	A21	NA					
	A24	113	0	0.92	1.21		4
	A25	77	11	0.81	1.32		2
	A33	50	0	0.95	2.11		0

[1] Improvement in PAT is not included

Complete exposure to the game (max. level) is evident for the group with good number processing skills. Group differences in exposure are evident. However, no such group difference is apparent in the number of numeracy tasks on which improvement is made. We observe that individuals from the poor and mixed groups may also have high exposure (A31, A34, A24) associated with good outcomes. Notable by contrast is the performance of A38, from the poor number processing group, who has the lowest exposure and the best outcomes of all. Perhaps surprising is the fact that mean score, supposedly sensitive to individual differences in number knowledge as well as gaming skills, does not distinguish between groups. A clearer relation between gameplay and number processing is observed in the more fine-grained measure of time per tile

Of those participants who did not show evidence of improvement, A23 was already a high performer in number processing before intervention and may therefore have been less likely to make progress. A33, who had a complex mixed profile, does not appear to have struggled with the game, working without repetition at a broadly average rate (2.11 seconds per tile). However, A33 completed relatively few levels (50), and made no progress on numeracy measures. A29, who appears to have struggled, working through 74 levels with 7 repetitions at a much slower rate (4.44 seconds per tile), also made no progress on numeracy. Finally, it is notable that A25, with 11 repetitions in 77 completed levels, worked at a relatively fast rate (1.32 seconds per tile), and improved on two numeracy measures.

A broad pattern emerges within which game exposure and rate of play are generally related to the initial state of number processing skills, but outcomes do not appear to be so determined. Of particular interest is the case of A38 where poor initial number processing and low game exposure are associated with superior outcomes, indicating that even limited gameplay may have positive effects.

6.4.5. Bonus Games

Apart from one participant, all PWA completed the bonus games as they arose. The post-it game was automatically triggered when the participant reached levels 20, 40, 50, 70, 90, 110 and 120, providing practice in recognition of a spoken number and selection of the corresponding Arabic numeral on one of the ‘letters’ for posting. Following a selection error, the chosen ‘letter’ was removed, so that an appropriate match would be made eventually. Stimuli were varied across levels, matching the number ranges being used in the main game.

A few participants took an average of up to 2.5 minutes to complete the post-it game. Most took less than a minute, and ‘posted’ the correct number as first choice. A few made as many as six requests for repetition of the auditory stimulus on average each time they played. Error rates were generally low, but a few individuals made as many as 10 errors per level on average.

The number-line game was automatically triggered when the participant reached levels 10, 30, 60, 80, 100 and 130, providing practice in Arabic number sequence completion corresponding to the number ranges being presented in the main game. Ten sequences were presented at each level, with a total of 40 trials (Arabic numerals to be appropriately placed) at each level.

Most participants were able to complete the majority of sequences, and to place most numbers without error.

Overall, the bonus games were successful in providing participants with practice in transcoding and Arabic numeral sequencing within their skill levels. A few individuals required repetition of spoken stimuli (in the post-it game) but were guided by the error-less learning algorithm towards success.

6.4.6. Individual trajectories

6.4.6.1 Participant A27

A27 is a 77-year-old female with an aphasia quotient of 71.1 (moderate). She enters the study with relatively good number processing skills.

At baseline she confused -teen and –ty items (e.g., writing 17 for target 70) and made errors with triple digits (e.g., selecting 5073 for target 573). Post-intervention she performed at ceiling on these tasks, making significant improvement in Number ID, Reading, Writing and also on Addition with carry.

On the functional assessment, she scored 12 pre-therapy, 14 post-therapy.

<u>Gameplay</u>

A27 finished the main game without having to repeat any levels. Her self-report indicates she completed this in 11 days of playing (6 hours 23 mins). She selected tiles relatively quickly, made relatively few errors and completed 95% of target sequences in full. Altogether she presents as an efficient player who understands and implements gameplay rules over the full 140 levels.

Detailed examination suggests that, where A27 does make errors, these reflect occasional faults in high level gameplay, rather than failure of sequence. That is, having started a sequence, she may explore options for a longer sequence, but fail to “deactivate” the sequence already selected. Most errors were made within the number ranges 10-15 and 50-70.

On the Post-It bonus game, which focusses on spoken to Arabic numeral transcoding, A27 completed all 7 games with high accuracy, making just 6 errors out of a possible 168. Five of her errors were expansion type errors (e.g., 370 for target 37, 2090 for target 209), similar to performance at baseline.

On the Number line bonus game, which focusses on the Arabic numeral sequence, she completed all 6 games. She made 14 errors in total, 12 of which were made in sequences involving ascending/descending 2s.

There are obvious consistencies between A27’s intervention experience (inputs from gameplay) and her outcomes. Teen-ty confusion and triple digit expansion are commonly found in transcoding. Both are targeted in the Post-It game, where the algorithm underlying player engagement is designed to support errorless learning. Overall exposure to the main game, as well as the Number Line game, both of which provide supportive cues for sequence production, may have helped A27 to improve in arithmetic. However, any suggestion that these elements are active ingredients in the intervention would be speculative, especially since A27 showed improvement on the control task as well as on numeracy tasks.

6.4.6.2 Participant A41

A41 is 63-year-old female with an aphasia quotient of 85.5 (mild). She has a mixed profile of number processing skills at baseline, with relative strengths in transcoding but poor counting skills.

Post-intervention there were significant improvements in number reading, counting forward and backward, and addition with and without carry. In the reading task, pre-intervention errors involved exclusively triple digit numbers. Triple digits were also problematic in the counting tasks, where they were either spelled out (e.g., 1-2-4 for 124) or omitted entirely. A41 also made errors with sequences in the middle range of numbers, and particularly at decade boundaries (59-60 in both directions, 81-80, 20-19, 11-10). Though errors occurred post-intervention, boundaries errors only occurred for 20-19 and 81-80 at P2, and production of sequences of triple numbers was greatly improved.

On the functional assessment, A41 scored 15 pre-therapy and 14 post-therapy.

Gameplay

In the main game A41 reached level 110 (the final ascending 3s level). She was quicker on the game than the average for the whole group, but made several errors per level.

A41 repeated the following levels: 43 (range: 52-78, ascending); 44 (range: 61-86, ascending); 100 (range: 2-52 even numbers, ascending). Errors on these levels may reflect a tile-press which had not registered in the game, as it often appeared that A41 had skipped ahead by 2 tiles.

Numerous inputs from the main game involved the decade boundaries which were problematic for A41 pre-therapy. The forward boundary 59-60 occurs in 18 sequences, on three of which A41 made an error.

A41 selected the backwards boundary 11-10 correctly on all 17 occasions when it occurred. On the boundary 20-19 she was correct 73 out of 74 times, on 60-59 correct 14 out of 15 times, and on 81-80 she was correct on all 24 occasions. It is worth bearing in mind that the boards on which these tiles are presented are arranged to constrain the decision process. Every consecutive choice is made within the limited set of adjacent tiles, in order to support fluent gameplay and thereby enhance sequence knowledge. Particular care was taken to provide opportunities for massed practice on decade boundaries.

Note that as observed with repeated levels, above, it is possible that errors result from technical failure to register individual tile presses. If that is the case, then A41 seems to be largely error-free with sequences which were problematic for her at baseline.

On the Post-It game A41 made just 4 errors, 2 of which are non-teen reversal (e.g., selecting 26 for target 62). On the Number line game, she made a total of 28 errors, 15 of which were in the 26-65 range. Nine of these errors involved a specific decade boundary (39-40-41).

As with A27, direct linkage between A41's intervention experiences and her outcomes cannot be directly established from current data, though it is possible that multiple opportunities to practise problematic sequences helped her to produce them more easily post therapy.

6.4.6.3 Participant A38

A38 is a 71-year-old male with an aphasia quotient of 61.5 (moderate), who performed poorly across all baseline assessments.

Errors were broad-ranging, some were teen-ty confusions (e.g., 40-14, 19-90, 380-318 in writing; 70-17 in forward counting; 90-19 in backward counting), some involved decade boundaries (e.g., 29-70, 60-56, 30-22 in forward counting; 59-55, 79-95 in backward counting) but there were many others less categorisable (e.g., 18-85, 36-33, 508-588 in writing; 7-5, 7-3, 38-36 in backward counting). A38 frequently required the stimulus to be repeated, and, in counting, was frequently assisted by writing on paper or in the air.

Post-intervention, there were significant improvements in number writing and counting forward. Though the strategies of writing on paper or in the air persisted post-intervention, A38 made fewer errors across sequences.

On the functional assessment, A38 scored 10 pre-therapy, 11 post-therapy.

<u>Gameplay</u>

A38 had the lowest exposure to the game of all participants reaching only level 40. Thus, he did not encounter any backward or skip count sequences. Despite his low maximum level his self-report indicated that he had played for the full 15 days specified (total time 4 hours, 11 minutes). Note that A38's mean score is high, but his mean time per tile is higher than the mean for the whole group (1.97), suggesting that he is slow at registering his response and moving through the sequence.

Though A38's rate of progress was slow, he took little advantage of the help facility (maximum use was 4 times per level). Most of his errors were made on teen sequences 10-15 and 10-20 (ascending).

Most problematic for A38, taking account of a range of indicators, was level 27, which comprises the number range 35-43 (ascending) presented in a 5x5 matrix. He scored only 77% of the maximum possible score, made 13 errors and spent, on average, 2.41 seconds per tile.

A38 presents a range of challenges to interpretation. He presented at assessment with some familiar patterns (teen-ty and decade boundary errors). His exposure to the game is limited, and he appears to struggle with sequence selection. His gameplay experience was relevant insofar as levels 27-40 focus on decade boundaries, and indeed he did show improved outcomes in these areas. However, many of his errors at assessment are difficult to classify, as are some aspects of his gameplay. Nonetheless, A38 entered the study with low levels of number processing skill and showed improvements attributable to intervention.

6.4.7 People with Aphasia: Summary

The assessments assembled for the study, and modified following Cycle 1, provided a coherent battery covering a range of numeracy components. The core measures of number processing proved to be highly inter-correlated, indicating a factor structure in which common representations of number symbols and sequences underpin the more complex procedures of calculation. Variability in levels of performance was extreme, causing ceiling effects on the one hand and floor effects on the other. Nonetheless the measures were sufficiently sensitive to reflect a range of contrasting patterns of performance classifiable as number processing groups, defined by performance in transcoding and counting tasks. The new functional numeracy test not only discriminated between number processing groups, but also showed significant moderate to high correlations with all other numeracy tasks, providing important evidence of content validity. The new Picture Analogy Test (PAT) avoided floor and ceiling effects and was, as intended, broadly dissociated from numeracy skills.

Of particular interest was the close association between aphasia severity and all numeracy tasks (including the functional test). More detailed analysis uncovered strong associations between object naming and number reading, and between auditory word processing and number identification. These findings suggest that, while basic number skills appear to form a unitary factor, they are nonetheless closely associated with variations in language skills. The association between measures was notably consistent across domains. However, the distribution of intervention outcomes did not reflect the same pattern. Improvements were found in both high and low number processing groups, and amongst those participants who showed a mixed profile. Attempts were made to understand this variation in outcome through examination of key parameters of gameplay. However, while there appeared to be an influence of exposure on outcome, and indications that the learning algorithms employed in the main game and the bonus games were effective, these effects were not consistent. In addition, there were exceptional individuals whose assessment and gameplay data proved hard to classify. There were also areas of uncertainty stemming from technical issues concerning data capture and processing.

Against the general pattern of association between numerical and linguistic skills, one individual shows a clear dissociation between number processing (tasks which draw on spoken number knowledge) and calculation. A22 enters the study with some preserved skills in calculation, but very low number processing levels. Following intervention his only area of improvement is in addition with carrying. There is no change in basic number knowledge, but the intervention appears to have improved his ability to carry out more demanding calculations involving carryover. It is possible that his working memory has been strengthened, and so he is better able to hold information in his mind and manipulate it.

6.5 Children with DLD: Intervention

6.5.1 Participants

Six children with DLD were recruited to the study. Four were recruited by contact with parents through E-DLD (Engaging with DLD: https://www.engage-dld.com/). Contact with these participants and their parents were carried out online via Zoom. Two participants were recruited through established links with their SLT and specialist teacher. These participants (GD25 and HD26) were given a reduced assessment battery, as required by staffing and timetabling constraints within the school day. Parents gave permission for language test data and reports to be made available to the research team. Table 8 shows age, gender and standard scores (i.e., scores corrected for age) on the Clinical Evaluation of Language Functions (CELF) (Wiig et al., 2013), where available.

Table 8. DLD participants: Age at baseline 1, gender and language status.

Participant	Age	Gender	CELF standard scores[1, 2]	
	years; months		Expressive	Receptive
AD20	5;9	male	4-6	7-9
BD21	7;1	female	3-5	1-6
ED23	7;2	female	4	7-8
FD24	7;10	female	4-7	5-9
GD25	7;7	male	3-4	2-6
HD26	6;8	female	Not formally tested	4-7

[1] Standard scores between 7 and 13 are within the average range.

[2] Where multiple subtests were administered and produced different scores, the range of scores is reported.

All participants were clinically diagnosed with DLD. FD24 had an additional diagnosis of 7q11.23 duplication syndrome (https://medlineplus.gov/genetics/condition/7q1123-duplication-syndrome/). Parental report indicated that speech delay was FD24’s primary symptom. AD20 was under investigation concerning possible coordination disorder. It was noted by the research team that AD20 had difficulty in maintaining attention during assessment, and that this was compounded by challenging technological problems. At times, successive online assessment trials took minutes to load, necessitating alternative means of communication (paper-based presentation and response via Zoom). Clinical reports on ED23 and GD25 also indicated significant attentional issues. HD26's expressive language had not been formally assessed. Clinical reports indicated that she had some difficulties with speech output, with expressive grammar and sentence structure. Both GD25 and HD26 were exposed to languages other than English at home.

6.5.2 Assessments.

The number processing and calculation assessment battery for children with DLD was identical to that used with PWA, with the addition of a simplified introductory section (Maths first level) for each of the calculation tasks. According to performance criteria participants would either stop after the first level task or proceed to the formal arithmetic tests given to the PWA. Number identification, number reading and calculation tasks were delivered online. The number writing task (completed by participants tested at home only) was accessed online and administered by parents, and so completed only by AD20, BD21 ED23 and FD24.

The functional numeracy test used with children with DLD was developed separately from the PWA version, with child-oriented content. Paper copies were sent to parents for completion once before intervention and once after intervention.

6.5.3 Pre-intervention status

Except for FD24, the children with DLD found online assessment challenging. AD20 and BD21, in particular, were at times unable to participate even with encouragement/support of their primary caregivers.

The majority of DLD scores fall within one standard deviation of the mean score for PWA. However, AD20, BD21 and GD25 all have numerous scores (including number processing, calculation and PAT scores) more than one standard deviation below the PWA mean. AD20 was the youngest participant at 5;9, was reported to have additional special needs, and suffered from major technical problems (extremely slow loading of successive online trials). For these reasons AD20 found the assessment battery, including PAT, challenging, with the notable exception of counting forwards (less subject to technical interference). FD24, on the other hand, performed at or close to ceiling level on all number processing tasks, with counting and calculation scores more than one standard deviation above the PWA mean. BD21, aged 7;1, some 16 months older than AD20, had more severe expressive and receptive language difficulties than AD20 relative to age, and found many of the assessments equally challenging. However, BD21 performed consistently and reasonably well on PAT, in striking contrast to her number processing scores. Likewise, GD25, aged 7;7, had some notably low number processing scores but performed consistently and reasonably well on PAT. Overall, the range of individual differences represented amongst the six participants is extremely broad. Notable also are inconsistencies between scores at baselines 1 and 2

Maintaining identity between assessments for PWA and children with DLD was successful in number processing tasks (transcoding and counting). However, in formal calculation tasks, the skills and experience of the child participants were often insufficient. This was anticipated, motivating the provision of a simplified pre-test. This was partially effective, supporting participation, but failing to provide a smooth transition to formal assessment, thereby setting an obstacle to reliable measurement of progress across pre-test and main assessment.

6.5.4 Intervention

Several constraints set limits not only to the size of the DLD sample, but also to the duration for which participants were available to the project. Following baseline measurements around one week apart, the SWAN game was made available to participants either at home or at school (GD25 and HD26 only) for a period of three weeks, with guidance that gaming sessions should last around 15 minutes per day. Those playing at home might choose to play at weekends if they wished. Following the conclusion of the intervention all participants were assessed a third time, but none was available for follow-up testing.

As in the case of participants with aphasia, the WEST approach to single case data analysis (Howard, Best & Nickels, 2015) was used to detect growth in proximal (number processing) or distal (calculation) skills. Howard, Best & Nickels (2015) provide a worked example of a repeated baseline design with single post-intervention measurement. Using this procedure, we can identify growth attributable to intervention but are unable to test for maintenance of any such growth.

Table 9 shows significant gains in assessment scores.

Table 9. Results showing significant gains in assessment scores.

Participant	PAT	Count fwds	Count bkwds	Number ID	Number Reading	Number Writing	Add No Carry	Add with Carry	Sub No Carry	Sub With Carry
AD20	NA			NA	NA	NA	NA	NA	NA	NA
BD21		✓	NA	NA	✓	NA	NA	NA	NA	NA
ED23					✓	✓	✓	✓	NA	NA
FD24	✓			✓		✓		NA	✓	✓
GD25		✓				Not tested				
HD26				✓	✓	Not tested				

NA: Not Available (either not attempted or missing through technology failure).

Note: ✓= Significant gain.

Empty cells: No evidence of gains.

FD24, who performed at a high level throughout the study, improved on PAT, as well as two measures of number processing. This pattern of results may indicate an improvement in domain-general cognitive skills (e.g., attention control). Only AD20, for whom engagement in assessment proved problematic, and whose recorded levels of performance were extremely low, failed to show improvement on at least one measure of numeracy. There is uncertainty concerning the reliability of AD20’s scores. BD21, also a low performer on assessment, improved on counting forwards and number reading. GD25, who also showed low baseline measurements, showed substantial improvement in counting forwards. Other participants showed improvement on two or three measures, largely limited to number processing tasks. Only ED23 showed improvement in calculation (addition without carrying). It should be noted that the main assessments of calculation, while allowing direct comparison with performance levels of PWA, made demands which often exceeded the skills and experience of the participants with DLD. Overall, despite the challenges presented by the assessment process, and the consequent extent of missing data, improvements in numeracy attributable to intervention were widespread.

The range of effect sizes observed for each task are shown in Table 10.

Table 10. Effect size range by task.

PAT	**Count fwd**	**Count bwd**	**No. ID**	**No. read**	**No. write**	**Add no carry**	**Add with carry**	**Sub no carry**	**Sub with carry**
-1 /+2.5	-15/11.5	-0.5/5.5	+0.5/+4	0/+5	2/+4.5	+4.5[1]	-1.0[1]	0.5[1]	-1.5[1]

While the data reported in Table 15 are sparse, the values for transcoding and counting register the breadth of individual differences in the effects of therapy, and the consequent importance of single case methodology. As with PWA, the PAT is notable for its stability.

6.5.5 Gameplay

Key gameplay parameters from the PWA sample, are shown in Table 11, alongside the corresponding values for each participant with DLD.

Table 11. Gameplay parameters for participants with DLD, compared to mean values for PWA.

	Maximum level reached	No. of repeated levels	Proportion of possible score	Mean time per tile	Mean no. of errors per level	Mean sequence proportion	No. of improved numeracy outcomes
PWA Mean (SD)	**96.39** (35.74)	**1.72** (3.01)	**0.91** (0.04)	**1.97** (0.91)	**6.56** (4.66)	**0.95** (0.02)	
AD20	22	9	0.74	3.05	91.23	0.86	0
BD21	29	19	0.73	2.31	221	0.85	2
ED23	28	22	0.72	1.32	210	0.83	3
FD24	130	8	0.90	1.15	17.3	0.94	2
GD25	22	27	0.70	0.70	79.2	0.83	1
HD26	15	45	0.74	1.20	255	0.86	2

FD24 completed 130/140 levels of the game performs within, or above, the mid-range for PWA throughout. It is worth noting that she is the oldest of the participants with DLD and has the least pervasive language difficulties. Her gameplay, in line with her assessment scores, distinguishes her absolutely from the remaining children with DLD.

All other participants perform far below the range of PWA, reaching levels between 15 and 29. AD20 has a slow rate of play, but relatively few repeated levels and low error rate. In this sense AD20 shows relatively skilled gameplay. BD21 also has a slow rate of play, but a higher proportion of repeated levels (19 repeated/29 completed) and high error rate. ED23 is a faster player but has similarly high rates of repetition (22 repeated/28 completed) and error. These data suggest that BD21 and ED23 are less skilled in gameplay than AD20. Both GD25 and HD26 have a strikingly high rate of repetition (27 repeated /22 completed and 45 repeated /15 completed) respectively. Error rate is high for HD26, but lower for GD25 than might be expected based on her repetition rate.

Notable is the relatively fast rate of play recorded by ED23, HD26, and especially GD25, whose mean time per tile is lower than any of the PWA. For all three participants, rates of play as well as error rates indicate an impulsive style of play. This accords with independent clinical reports of impulsivity for ED23 and GD25.

Except for FD24 (and arguably of AD20), the gameplay characteristics of children with DLD indicate a low level of competence and a possible failure to understand or to comply with the basic rules of the game concerning sequence entry. In the cases of GD25 and HD26, the players fail to reach criterion for progression more often than they succeed.

As noted for PWA, while gameplay characteristics discriminate clearly between players with DLD, with FD24 far more skilled than the rest, and GD25 and HD26 appearing to be far less skillful, there is no indication that skill level is associated with improved outcomes.

6.5.6 Bonus games.

Most participants only reached level 20, and so played the Post It game just once. AD20 engaged with the game, playing at a relatively slow rate, perhaps indicating concentration. He clicked to repeat the spoken on three occasions, made errors, but completed the game, perhaps dependent on the errorless learning algorithm. BD21 also played at a relatively slow rate. She made errors, did not request repetition of the stimuli, but completed the game. ED23 and GD25 played faster and were largely error-free. FD24 completed all levels, played at a relatively fast rate and with high accuracy.

It appears that the Number Line proved more challenging than the Post It game, particularly for AD20 and ED23, who may have chosen not to participate. GD25 completed the game at level 10, but often failed to complete the sequences presented. It is possible that AD20, ED23 and GD25 lacked the coordination skills to engage fully with the game. BD21 and HD26, by contrast, managed to complete more than half the sequence, making a similar proportion of correct placements on first try. As with the Post-It game, FD24 completed all levels and performed competently throughout.

Based on the sparse data available it appears that the Post It game, with its emphasis on transcoding, was accessible and potentially valuable for participants, some of whom may have been challenged by the co-ordination requirements of the number-line game.

6.5.7 Individual trajectories

6.5.7.1 BD21

BD21 entered the study with low levels of skill in transcoding and counting. She made progress attributable to intervention in number reading and in counting forwards. Here we examine in detail the challenges she met in playing the SWAN game, and how she dealt with those challenges. Overall, she made 19 repetitions during her play from level 1 to level 29.

BD21's gameplay behavior is progressive in a number of ways. She maintains a relatively high proportion of error free sequence selections despite increases in demand (range, sequence type, board size). When she fails to reach criterion for progression, she is persistent and is usually able to increase her score with two or three further attempts.

Of interest is the fact that BD21 shows systematic improvement through the repetition of levels 18 and 20, which focus on numbers in the teens. Her reading of numbers in this range is significantly enhanced in Post Therapy assessment. Similarly, she improves her scores substantially through repetition of levels 21 and 23 (counting in tens), and, in level 23, making use of help provided by the game. Though this aspect of counting skill was not included in formal assessment, specific improvement in this area was highlighted by parental report. General improvement in counting forwards was confirmed by formal assessment.

Parent report states (a) BD21 enjoyed the game and wished to continue after the end of the therapy period; (b) the introduction page at the start of each level helped her to understand what to do; (c) the game helped her to learn numbers 30, 40, 50 etc.

There is convergent evidence that SWAN helped BD21 to increase her number knowledge.

6.5.7.2 GD25

GD25 was identified with pervasive expressive and receptive language difficulties, and with attentional difficulties affecting his learning. He entered the study with low scores in transcoding and counting. In counting forwards, he had problems entering the number sequence at points beyond one. This phenomenon of an 'unbreakable sequence', reported by Fuson (1988) is widely observed, but rare in a child aged 7;7. GD25 also had difficulties producing reliably distinctive forms of teen and decade numbers. Following intervention, he was able to enter the number sequence after one, and at later stages in the teens and decades. He was also able to produce distinctive forms of teen and decade numbers.

As noted in 6.5.4, GD25 presents as an unskilled and impulsive gameplayer, with a high repetition rate. Observing his overall progress from level 1 to level 22, we observe that 9 of his 27 repetitions take place between levels 1 and 10, which deal with sequences in the range of numbers 1-10, all starting at 1. Levels 11-15 are also devoted to sequences up to 10, but require entry at 2, 3, 4, or 5. It is here that GD25's repetitions are most numerous. 15 times he was required to repeat these levels. At level 14 alone he was required to repeat 8 times, and to enter no less than 93 sequences. Of importance is the fact that he waited for the help provided by the game, and implemented that help, on average 6.16 times in each of those 93 sequences. Eventually GD25 succeeded, achieving 79% of the maximum possible score for the level, thereby reaching criterion and progressing to level 15. It is possible that the help system was effective in maintaining GD25's motivation during this prolonged period of repetition.

Between levels 16 and 22, GD25 has only two repetitions, quickly adapting to the requirement to enter teen sequences starting from 10 and 11, and then to select sequences in tens up to 50. The post-it bonus game (after level 20) gave GD25 further exposure to teen/ty contrasts. On one trial he misidentified 'thirteen' for 31, and 30. The errorless learning algorithm then allowed him to make the appropriate match.

As we saw in the case of BD21, so with GD25 there is clear association between specific content and characteristics of gameplay, and specific improvements in outcome. Thus, to dismiss GD25 as an unskilled player of the SWAN game, may be to miss the point that GD25 appears to have taken from the game what was specifically necessary for him to enhance his number knowledge.

Informal feedback from professionals working with GD25 was that he was engaged by the game, took advantage of the tutorial information provided at the beginning of each level, and was extremely independent. He and his fellow participant would remind the teacher when it was time to play and would maintain good concentration during the session, even though attentional difficulties affecting his learning had been reported.

6.5.8 Functional Numeracy Test

A newly developed functional numeracy test was delivered in paper form to participating families, with the request that they be administered once before therapy and once afterwards. Results are available for three participants.

AD20 engaged with the test and improved significantly on repeated assessment. Five out of 9 correct answers pre-therapy were for questions requiring simple number reading. Post-therapy, the improved score included two calculation questions, and two questions requiring AD20 to read time aloud. Without a repeated baseline measurement, we do not attribute this change to intervention, but simply report that it took place, possibly as a result of practice. Particularly interesting are the relative performance levels of ED23 and FD24. On most numeracy measures FD24 showed skills far in advance of ED23, while on functional skills they are closely matched. The possibility that a discrepancy of this sort could occur more widely should be investigated in further studies.

6.5.9 Children with DLD: Summary.

The six children with DLD recruited to the SWAN study comprise substantial individual differences. FD24, the oldest participant, presents a different communication profile from the rest, with primary problems in speech output reflecting her additional genetic diagnosis of 7q11.23 deletion syndrome. Compared to the rest of the children, she performs well in all numeracy assessments, and makes excellent progress through the SWAN game. Following intervention she showed improvement on the non-numeric control task on PAT, as well as two measures of number processing. This pattern of results may indicate an improvement in domain-general cognitive skills (e.g., attention control). AD20, the youngest participant, was unfortunately subject to major technical problems in assessment, and was unable to provide reliable assessment data. Nonetheless, based on gameplay data, AD20's engagement with the SWAN game appears to have been successful. His rate of play was slow, but he made progress through the game with relatively low repetition and error rates. BD21, ED23, GD25 and HD26 all entered the study with low numeracy scores and showed significant improvement on one or more of these measures post-intervention, while showing stable performance on the non-numeric control task. Detailed investigation of individual gameplay characteristics (BD21 and GD25) demonstrates clear associations between exposure to specific areas of game content and specific improvements in assessment scores. The case of GD25 is particularly interesting. High repetition and error rates would seem to indicate low competence in gameplay, and consequent low expectation for intervention-based improvement. However, GD25 confounds this expectation, appearing to benefit from the learning support provided by the game and from the extended exposure that his gameplay style entails.

Maintaining identity between assessments for PWA and children with DLD was successful in number processing tasks (transcoding and counting). However, in formal calculation tasks, the skills and experience of the child participants were often insufficient. The provision of a simplified pre-test was partially effective but failed to provide a smooth transition to formal assessment, thereby setting an obstacle to reliable measurement of progress. Thus, evaluation of the proximal (number processing) vs. distal (calculation) effects of intervention is limited by measurement issues. Nonetheless, the case of ED23, who was able to access the formal assessments, provides evidence that distal effects are possible.

Despite the small sample, results from children with DLD provide indicative evidence that the SWAN game can feasibly be delivered in home and school contexts and is capable of delivering significant gains in basic numeracy.

Informal feedback confirms these findings. Two teachers (one a parent of a participant) and an SLT working with the SWAN project team have strongly endorsed the game. Both suggested that it would work as an effective tool for small group work in the classroom, either in a DLD unit, or else in mainstream where it could serve as an effective support for lower performing groups in Year 2. The game was seen not only as supporting number skills, but also attention, visual orientation, hand-eye coordination. Both teachers were interested in the possibility of real time reporting of progress to the teacher's laptop. The SLT we worked with would like to see the game used for a daily 15-minute period for longer than the three weeks specified in the SWAN intervention. Further observations were made concerning the effectiveness of the game in engaging two children identified with attentional deficits. Both children, once provided with the tablet and basic instruction, were able to concentrate on the game in future sessions without any additional management. Both were sad to learn that the tablets had to be returned on completion of the project.

7. Discussion

The aim of the project was to evaluate the feasibility of the SWAN game as an intervention to support basic numeracy in PWA and children with DLD. Challenges presented by the COVID-19 pandemic were met by developing on-line assessments of initial state of numeracy and post-intervention outcome. For PWA this proved successful insofar as the recruitment target of 15 self-selected single cases was exceeded (18 completed the study). Data from PWA revealed variations in the presentation of numeracy deficits which were readily categorized as subgroups which may prove valuable as the identification of the unmet clinical need of numeracy support becomes a priority in the field. A coherent factor, structure of individual differences in basic number processing skills, associated with more complex calculation procedures, was observed. Detailed correlational findings indicate a close mapping between patterns of numeracy deficit and language profiles, though outstanding exceptions are noted. Widespread improvements in both number processing and calculation, attributable to intervention, were observed. These were not limited to particular subgroups of participants but occurred across the range of pre-intervention skill levels. Within the group of PWA, gameplay characteristics such as high exposure (number of levels played) and fast rate of play (time per tile) were associated with high pre-intervention numeracy skills, but also occurred often in the low and mixed skill groups. Intervention-based improvements in outcome were observed in participants whose gameplay competence appeared low, as well as in skilled players.

Despite the low sample size, and the challenges of online assessment for children with DLD, widespread improvements attributable to intervention were observed in the immediate post-intervention period. Compared to the performance levels of PWA, children with DLD were, for the most part, more limited not only in numeracy levels, but also in gameplay skills. However, detailed examination of individual gameplay content and characteristics suggested that significant improvements in outcome may occur in the context of apparently low gameplay skills. There is an indication the specific gaming features designed to support learning may have been influential across the range of both numeracy and gaming skills. Notable also is the finding that participants whose attentional skills were cause for concern, and whose gameplay was impulsive, were nonetheless able to concentrate on the game independently and made numeracy gains attributable to intervention.

The effects of COVID-19 on the SWAN project were profound. The move to on-line and remote delivery of all assessments presented major obstacles, unanticipated in the original project design. These issues, alongside funding constraints, proved to be particularly frustrating in relation to the recruitment of participants with DLD. The final sample of six participants fell far below expectation.

Other important limitations were noted regarding gaming technology. There were, and continue to be, areas of uncertainty about data capture. While the gameplay data reported above are deemed reliable, there are other variables whose validity and reliability remain in question. These issues should be resolved in any future study.

8. Conclusion

The study has provided strong indications that the SWAN game may be used as an intervention to enhance basic numeracy in PWA. Similar indications are found for children with DLD, though the data are sparse. The study has confirmed the importance of unmet needs for numeracy support in PWA and provided a possible framework for categorization of such needs. In both PWA and children with DLD, improvements attributable to intervention were observed across the range of pre-intervention numeracy levels and gameplay skills. There are indications that the learning algorithms written into the SWAN game were successful in supporting participants with a wide range of learning difficulties. Technical issues concerning data capture should be resolved in any future studies.

Further data needs to be collected to establish whether these findings can be replicated or indeed surpassed when the intervention is being delivered in less challenging conditions. We still need to determine what is an ideal gaming experience – time, using headphones to limit distractions, social engagement etc.

It is recommended that future research focuses on developing the SWAN game so that it targets the needs of different client groups more effectively. Some children with DLD appear to need a slower progression and more support in the early stages of the game; while PWA, might benefit from a wider range of bonus games to support targeted computational skills and encourage progress.

Of great value would be the development of algorithms that would allow players to progress through the game according to their abilities. Individuals with more advanced skills could enter the game at a higher level and focus more on challenging number sequences.

Other adaptations to the game, including voice recognition, would allow us to investigate whether both output and input processing need to be targeted if individuals are to make maximal gains.

The SWAN project has generated interest amongst researchers in the field as well as colleagues in health and educational settings. Future work could also include trials with individuals with poor number skills who do not have language problems.

References

Brookman-Byrne, A., Mareschal, D., Tolmie, A., Dumontheil, I. (2019). The unique contributions of verbal analogical reasoning and nonverbal reasoning to science and maths problem-solving in adolescence. *Mind, Brain and Education*, 13, 211-223.

De Luccia, G., Ortiz, K. (2016) Ability of aphasic individuals to perform numerical processing and calculation tasks. *Arquivos de Neuro Psiquiatrica*, 72, 197-202.

Donlan,C., Cowan,R., Newton,E.J., Lloyd,D. (2007). The role of language in mathematical development; Evidence from children with specific language impairments. *Cognition* 103, 23-33.

Durkin, K., Shire, B., Riem, R., Crowther, C., Rutter, D. (1986). The social and linguistic context of early number word use. *British Journal of Developmental Psychology*, 4, 269-288.

Fuson, K.C. (1988) *Children's counting and concepts of number*. New York: Springer.

Habermann, S., Donlan, C., Göbel, S.M., Hulme, C. (2020) The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development. *Journal of Experimental Child Psychology,* 193, 104794.

Howard, D., Best, W., Nickels, L. (2015) Optimising the design of intervention studies: critiques and ways forward. *Aphasiology*. 16, 151-168.

Hulme, C., Brigstocke, S., Moll, K. (2016). *Test of Basic Arithmetic Skills*. Oxford University Press.

Kertesz, A. (2007) *The Western Aphasia Battery-Revised*. San Antonio, Texas. PsychCorp.

Korpipaa, H., Moll, K., Aunola, K., Toivainen, A., Koponen, T., Aro, M., Lerkannen, M. (2020) Early cognitive profiles predicting reading and arithmetic skills in grades 1 and 7. *Contemporary Educational Psychology, 60,* 101830.

Malone, S., Burgoyne, K., Hulme, C. (2020) Number knowledge and the approximate number system are two critical foundations of early arithmetic development. *Journal of Educational Psychology*. 112, 1167-1182.

Morris, K., Ferguson, A., Worrall, L. (2014*) International Journal of Speech-Language Pathology.* 16, 541-551.

Middleton, E., Rawson, K., Verkuile, J. (2019) Retrieval practice and spacing effects in multi-session treatment of naming impairment in aphasia. *Cortex*, 119, 386-400.

Siegler, R.S. (2016). Magnitude knowledge: the common core of numerical development. *Developmental Science*. 21, 341-361.

Varley, R., Klessinger, N., Romanowski, C., Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences of the USA*, 102, 3519-3524.

Wiig, E., Semel, E. & Secord, W. (2013) *Clinical Evaluations of Language Fundamentals - 5*. London: Pearson.

Yuan, L., Prather, R.W., Mix, K.S., Smith, L.B.(2019) Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition*, 189, 89-104.