# Orchestrating Numeracy and the Executive - "The ONE" Programme

End of Grant Report for the Nuffield Foundation

Preprint Date: 12th March 2023

Authors

Gaia Scerif[1*], Sylvia Gattas[1], Amy Godfrey[1] Zachary Hawes[2], Steven Howard[3], Rebecca Merkley[4], Rosemary O'Connor[1] and Jelena Sučević[1]

[1*]Department of Experimental Psychology, University of Oxford, UK

[2] Department of Applied Psychology and Human Development, University of Toronto, Canada

[3] School of Education, University of Wollongong, Australia

[4] Department of Cognitive Science, Carleton University, Canada

[*]Corresponding Author:

Gaia Scerif, Professor of Developmental Cognitive Neuroscience

Attention, Brain and Cognitive Development Group, Department of Experimental Psychology, University of Oxford, United Kingdom

E-mail: gaia.scerif@psy.ox.ac.uk ORCID: 0000-0002-6371-8875

The first author acknowledges responsibility for framing this overall report. All other authors are listed alphabetically in recognition of diverse but equally important contributions.

**Acknowledgements**

A Project Grant by the Nuffield Foundation (to Scerif [PI], Hawes, Howard, and Merkley, "Fostering resilience by injecting executive challenge into early mathematics") supported this project. 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, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. www.nuffieldfoundation.org

We are very grateful to all Advisory Board members for the Project Grant, for their intellectual contributions during advisory board meetings and beyond: Emma Blakey, Jennie Challender, Aleisha Clarke, Keely Cook, Katy Jeary, Ruth Maisey, Gill Mason, Joanne Mason, Edward Melhuish, Fionnuala O'Reilly, Kathy Sylva, Ellen Wright and Victoria Simms. We are also heavily indebted to Hannah Andrews, Angelina Bogdanova, Abigail Heath, Libby Kent and Francesca Plaskett for contributing to post-intervention data collection. Finally, none of this work could have been achieved without the huge support and effort of children, early years educators and parents at our volunteering settings.

# Table of Contents

# Executive Summary

The current study focused on the collaboration between cognitive scientists and educators to co-develop and progressively refine the Orchestrating Numeracy and The Executive ("The ONE") Programme, an evidence-based integrated Executive Functions and Mathematics intervention composed of professional development and play-based activities. This iterative process resulted in high ratings of acceptability and feasibility, as well as acceptable adherence and fidelity across settings, combined with knowledge gains for early years practitioners and preliminary evidence of improved early numeracy, in particular for disadvantaged children. At the same time, the study highlighted general and specific barriers to intervention implementation in early years settings, and multiple possible steps for future refinement. Given the initial preliminary evidence, the intervention is being further evaluated in a larger scale trial funded by the Education Endowment Foundation and The Stronger Practice Hubs.

## Context and aims

Executive functions (EFs) are thinking skills that support our ability to maintain information in memory, inhibit inappropriate actions and think flexibly. Evidence points to an interplay between early mathematical skills and executive functions (e.g., Coolen et al., 2021), with executive abilities predicting mathematical skills in preschoolers. However, many children from disadvantaged socio-economic status (SES) backgrounds may live in environments that offer fewer opportunities to develop both EFs and foundational mathematical skills (Blakey et al., 2020; Hanner et al., 2019; Raven and Blair, 2004).

There is extensive evidence that early years mathematics and executive functions interplay dynamically (e.g., Coolen et al., 2021), but interventions that have focused exclusively on executive functions in isolation have failed to result in improvements in mathematics, suggesting that a different approach is needed to facilitate transfer of EF improvements onto mathematics (Scerif et al., under review). The ONE programme aims to encourage early years educators to embed executive function demands into mathematics activities, to challenge the development of children's early mathematical skills. The programme consists of a combination of professional development (PD) sessions and a set of 25 mathematical activities with embedded EF challenge for practitioners to play with children over the course of 12 weeks. A primary outcome of this study was to gather evidence on whether the intervention is acceptable and feasible for practitioners, allowing for integration into their routines. As a secondary outcome, the study aimed to preliminarily test whether mathematical outcomes and executive functions would improve more in children receiving the intervention compared to business as usual.

## Methodology

*Setting-level:* Four settings took part in the co-development stage and an additional fifteen settings took part in the main study (N = 6 local authority maintained, N = 9 private or voluntary). As a measure of setting quality, observations took place using the Sustained and Shared Thinking and Emotional Wellbeing Scale (SSTEW, Siraj et al., 2015) with an author-developed added mathematics subscale (O'Connor et al., in prep).

*Practitioner-level:* Practitioners completed questionnaires on their prior experience, attitudes towards teaching and mathematics, and classroom behavior (N= 51 questionnaires, out of approximately 75, returned in total). Practitioners varied in terms of qualifications and experience both between and within preschools.

*Child-level:* 193 children took part in the study (mean age = 47.2 months, 41 - 54 months). Each child completed a combination of standardized and experimental but previously used assessments (seven mathematics and four EF assessments) at two timepoints – before (T1) and after (T2) the intervention. Post-intervention assessors were blind to the intervention group. Following EF literature in this age group, the four EF assessments were combined, producing a single latent measure of EF. Parental and practitioner questionnaires were used to gather data on socio-economic status (SES), the home environment and the child's behaviour. Child eligibility for early years pupil premium status (EYPP) as reported by the preschool or the parents was used as a proxy for low SES, with eligibility equating roughly to a family income of less than £16,500 per year.

*The Intervention:* Eight preschools were pseudo-randomly assigned to the intervention group and a further seven preschools formed a business-as-usual (BAU) control group by a team member who was blind to all setting-level information except overall setting postcode multiple index of deprivation (< or = 5, >5), size (large, medium, small), location (urban, rural), and type (LA-maintained, PVI). Practitioners at each setting were asked to attend four PD sessions, each spaced one week apart, and play three activities per week with the children over a 12-week period. They were given the opportunity to discuss, ask for advice, and provide feedback during PD sessions, in a follow-up phone call in week 8 and in a final interview/observation session in week 12. As a measure of adherence, practitioners were asked to record all activities on a poster. Post-intervention observations and interviews were used as measures of fidelity and quality of delivery.

## Primary Outcomes

Of the eight intervention preschools, five settings adhered to delivering more than 75% of activities, ranging from 100% to 22.5% of the recommended 3 activities per week over 12 weeks. An implementation quality score was calculated for each preschool using several indicators from the observations, interviews, activity posters and evaluation sheets. Mind maps revealed quantitative knowledge gains related to executive functions and qualitative changes in knowledge related to both executive functions and mathematics. The intervention was generally found to be acceptable and beneficial. Much of the feedback in interviews was positive, indicating that practitioners felt that they had benefited from the experience, gaining both knowledge and inspiration. However, qualitative data also revealed complex barriers to taking part, including barriers related to the intervention (e.g., difficulty of activities compared to children's level of ability), and external factors (e.g., staff shortages, staff turn-over and therefore difficulties with engagement).

## Secondary Outcomes and Measure Development

Post-intervention and after controlling for differences at time 1, the intervention group achieved higher scores than the BAU group on the Early Years Toolbox Numbers (EYTN) scale, a combined numeracy measure, and on Corsi Blocks, a short-term memory scale. Modelling EYPP eligibility revealed interaction effects of intervention and EYPP: EYPP-eligible children in the intervention group scored higher than those in the BAU group on three numerical measures (EYTN, a number comparison task, and a spatial task, the British Ability Scale Pattern Construction subscale, BAS PC), on executive functions overall, and on two measures of short-term memory (Corsi Blocks and Mr Ant). Children's attitudes to early numeracy, measured via a child-friendly author developed self-report scale, correlated with their numeracy performance, inhibitory control and working memory skills.

# Orchestrating Numeracy and the Executive – The ONE project

## Introduction

Current evidence highlights the interplay between early mathematics learning and executive functions (e.g., Coolen et al, 2021). Executive functions are cognitive skills that help us maintain goals in mind, inhibit inappropriate responses and think flexibly. Research suggests that executive skills predict mathematics skills and their growth for four-year-olds prior to school entry, although these rarely feature in Early Years practitioner training. Socio-economically disadvantaged children are increasingly understood to have fewer opportunities to practice executive functions in their everyday environments (e.g., Raver and Blair, 2014) and fewer opportunities for extensive every day early mathematics exposure (e.g., Hanner et al, 2019). This suggests a vicious cycle of poor exposure and practice for these two inter-related skills. Recent empirical data, therefore, strongly suggest that successful mathematical learning requires the integration of mathematics-specific skills and executive functions (Scerif et al., under review).

The aim of the Orchestrating Numeracy and the Executive (The ONE) programme is to support Early Years practitioners to run play-based mathematics activities that in turn foster mathematics development by embedding executive function skills into mathematics learning. The programme consists of professional development (PD) for practitioners, a pack of 25 activity cards, and some low-cost resources to be used with the activities. The cards highlight how to identify and gradually increase executive function demands within mathematics learning, so that practitioners can scaffold their pupils' mathematics learning at the optimal level of challenge, with the aim to boost early mathematics development. The activities last five to ten minutes and can be embedded into preschool routines such as small group activities, outdoor play, and free play. Practitioners are asked to engage in a minimum of three activities per week for the 12-week duration of The ONE.

A predecessor programme integrating executive challenge into play-based activities (without the mathematics focus) was trialed in Australia (PRSIST, Howard et al., 2018, 2020). This programme implemented a similar schedule of PD and frequency of play-based activities to The ONE. It resulted in improvements in executive functions for the intervention settings, though improvements in attainment did not reach statistical significance. The ONE adapted the Australian programme to engage mathematics-specific content, given mounting evidence that executive functions are key for early mathematical development. This content was co-developed by cognitive scientists, with early years teachers and practitioners in pilot settings. The co-developed intervention has been piloted in the UK and underwent a small scale feasibility randomized controlled trial in 2022. The acceptability, feasibility, adherence, fidelity and preliminary small scale efficacy findings are summarized here.

# Methods

## Participants.

**Setting-level demographics.** Four preschools local to Oxford took part in the co-development of intervention materials and sixteen preschool settings from Oxford and the surrounding areas volunteered to take part in the efficacy phase of the study, with one setting withdrawing due to COVID pressures (final M = 12.9 children per setting, SD = 3.75). Settings were pseudo-randomly allocated to the intervention group, with efforts made to match the groups on Indices of Multiple Deprivation (IMD) decile and Income Deprivation Affecting Children Index (IDACI) based on the postcode of the preschool, setting size (small/medium/large), setting location (urban/rural) and setting type (private/charity/local authority funded) resulting in 8 settings in the intervention group and 7 in the control group. Settings were well matched for IMD ($M_{intervention}$ = 5.5, SD = 2.14 ; $M_{control}$ = 5.5, SD = 2.62), Income Deprivation Affecting Children Index (IDACI) ($M_{intervention}$ = 5.0, SD = 2.82 ; $M_{control}$ = 5.0, SD = 2.39), setting size ($M_{intervention}$ = 30.9 children, SD = 21.0 ; $M_{control}$ = 30.7, SD = 25.5), and setting type (intervention = 42.9% private; control = 62.5% private). As a measure of quality of staff-child interactions, the Sustained Shared Thinking and Emotional Well-being scale (SSTEW; Siraj et al., 2015) was used. This was supplemented by a new mathematical sub-scale (see appendix tables A1-A4), which was designed to reflect areas of mathematics in the Early Years Foundation stage (EYFS) guidelines. These measures revealed a high level of variability in performance, with preschools globally performing well on scales related to language and social development, but with lower scores on scales related to critical thinking and mathematics (see Figure 1).

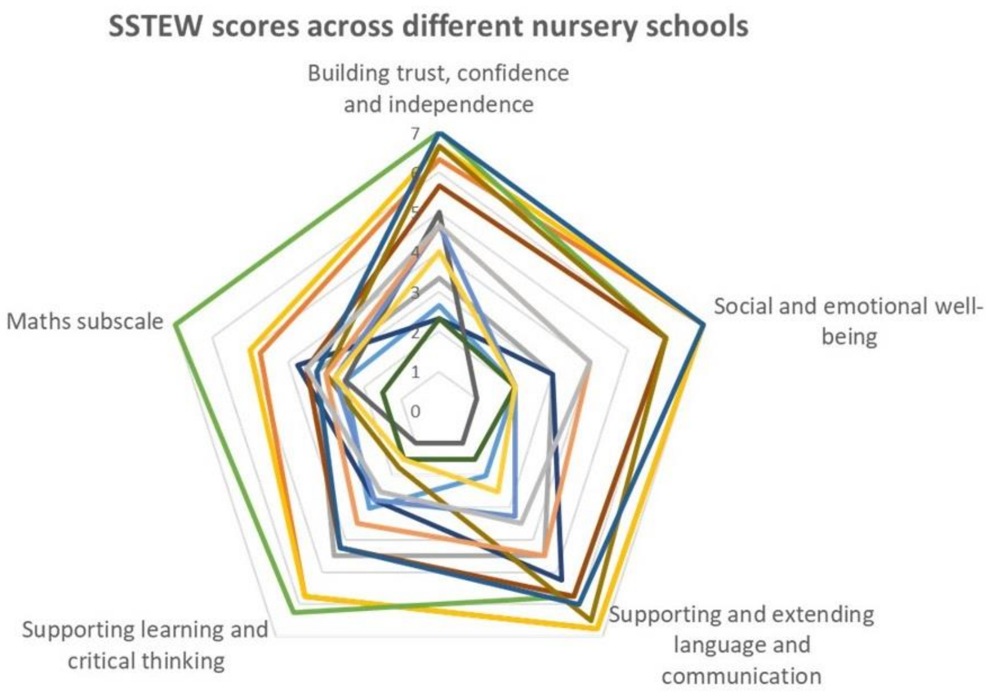



*Figure 1: Summary of SSTEW subscale scores across settings (with colours representing each individual setting).*

**Practitioner-level demographics.** Paper and online copies of a practitioner questionnaire were given out to practitioners at all preschools. 51 questionnaires were returned, with a mean of 3.4 questionnaires returned per setting and at least one questionnaire submitted at each setting. 49% of practitioners had more than 10 years of experience working in an Early Years environment, whilst 17.6% had been working in early years for less than a year. 76.4% of the practitioners who returned the questionnaire reported having some form of early years qualification, of which 23.5% of practitioners held a level 6 qualification, which is equivalent to fully qualified teacher status or a postgraduate degree in child development and 49.0% held a level 2 or level 3 qualification – equivalent to a post-secondary diploma.

**Child-level demographics.** The full study sample consisted of 193 children (M age = 47.2 months, range = 41-54 months, 111 female) who were pseudo-randomly allocated by setting to the Business-As-Usual (BAU) control group (N = 90) and the Intervention group (N = 103), whose key demographics are reported in Table 1. A combined variable was created for Early Years Pupil Premium (EYPP) eligibility based upon reporting at the nursery school level (N=147) and parent-reported income (N=77), as indicators of EYPP eligibility. Of the 161 children for which either nursery-reported or parent-reported eligibility was available, 24.8% were deemed to be eligible for the fund. The Brief Early Skills and Support Index ***(BESSI,*** Hughes et al., 2015) was used as a measure of school readiness. This is a questionnaire completed by the practitioner and divided into 4 subscales: (1) behavioural adjustment, (2) language and cognition, (3) daily living skills, and (4) family support. Each item is given a score of 1 (strongly agree or agree) or 2 (strongly disagree or disagree), with a higher score representing more problem behaviours. This scale has been validated on a sample of 1451 children in the North West of England (Hughes et al., 2015), in which each of the 4 sub-scales was found to have strong internal consistency. The four subscales have been combined to make a single score. In addition to our pre-intervention child-level assessments (described below), it provided us with an overall comparator for the two groups of children.

**Table 1. Summary of the demographic information for the control and the intervention group**

| **Measure** | **BAU Control group** | **Intervention group** |
|---|---|---|
| Number of participants | 90 | 103 |
| Age (months) | 47.2 | 47.33 |
| Sex (female, male) | 51, 39 | 60, 43 |
| EYPP eligibility (%) | 15 (21%) | 25 (28%) |
| Special educational needs (SEN) | 6 | 5 |
| English at home | 72 | 70 |
| Another language | 22 | 38 |
| BESSI score | 1.16 | 1.19 |

**Note.** For children for whom information was not returned (e.g., English at home was not reported), the data were treated as missing (rather than imputing values), which accounts for variable totals across demographic characteristics.

## Procedure.

### Pre and Post-Intervention Child-level Assessments.

All children were tested individually in a quiet space within nurseries. Each child completed brief tasks across two sessions, with each session taking approximately 30 minutes, carried out on two separate days, with sessions counterbalanced across children. Task order was chosen to ensure a balance of manipulative and iPad-based tasks and to avoid starting with tasks that require the child to speak. Post-intervention assessments were carried out by researchers who were blind to trial arm allocation.

***Numeracy. Overall numeracy (Early years toolbox; Howard et al., 2021)***. The early years toolbox numeracy (EYTN) task is a tablet-based measure of general numeracy skills. Children responded to a series of problems presented on the iPad by either tapping the screen (e.g. “Tap the largest tree”) or providing a verbal response, which the experimenter then records in the app (e.g. “How many cats are there?”). Items on the task are split into various different mathematical domains: number sense, cardinality and counting, numerical operations, spatial and measurement constructs and patterning. The game ended after 5 consecutive mistakes and on average took 8 minutes. The total accuracy score was used for analysis, with one point scored for each correct item. The EYTN was validated on a set of 246 Australian 3-5 year-olds, and showed good test-retest reliability (r (46) = .89) (Howard et al., 2021). ***Component numerical skills (symbolic and spatial skills). Count High (Coolen et al., 2021).*** To assess children's counting skills, children were instructed to count as high as they could and the highest number reached without having made any mistakes was recorded, stopping at 100 If the child was able. This task has previously been undertaken with British preschoolers (Coolen et al., 2021), when it showed good sensitivity to change across a 5-month period. ***Digit Comparison (adapted from Nosworthy et al., 2013).*** This task is designed to measure children’s symbolic number comparison abilities. Two number digits (1-9) were presented side by side on the screen of a tablet and the child was asked to tap the larger of the two numbers. The final score was calculated as a proportion of the number of items answered within 1 minute. ***BAS3 Pattern Construction.*** The pattern construction scale from the third edition of the British Ability Scale (BAS3; **Elliott, & Smith, 2011**), was used as a measure of spatial ability. This scale requires children to copy patterns using wooden blocks, foam squares and plastic cubes with different patterned and coloured sides. The BAS3 was standardised on a sample of British children including 269 3- or 4-year-olds. The reported corrected Rasch split-half reliability for the pattern construction scale was .89 (Elliott & Smith, 2011). ***Give N (adapted from Cahoon et al., 2021).*** A version of the Give-N task (Wynn, 1990) was used as a measure of cardinality, following the adapted procedure outlined by Cahoon et al. (2021). Children were asked to place a given number of plastic fruit on a plate for 3 blocks of 5 trials, using numbers 3, 4, 6, 11 and 15. Once the child had placed the items on the plate, the researcher asked “Is that [n]?”. If the child responded “Yes”, the researcher proceeded to the next trial. If the child responded “no”, the researcher repeated the original request. For higher numbers (6, 11 and 15), the child was asked “Can you count and make sure it’s [n]?”. The final score was the number of correct items out of a possible 15. On a similar task, Batchelor et al. (2015) reported Cronbach’s alpha of .76. ***Number naming (Nosworthy et al., 2013).*** As a measure of symbolic number knowledge, children were presented with a laminated sheet containing each digit from 1-9 twice in a random order, resulting in 18 total digits. The researcher pointed at each digit in turn, asking the child, “What number is this?”. The score used was the number of correct items out of a possible 18. ***Order Processing (Coolen et al., 2021).*** Children were presented with a set of three number cards, each containing one Arabic numeral (1-9), which they were asked to place in order from smallest to biggest. These three numbers were either sequential (1,2,3), with a gap of one (1,3,5) or with a gap of two (1,4,7). Following 4 practice trials, there were 12 main trails. The task ended after six cumulative mistakes. A total score out of 12 was calculated for analysis**. *Attitudes to Early Numeracy (Child Attitudes Questionnaire – Preschool, Gattas et al. in prep*)** consisted of 23 items asking children to report on their emotions and attitudes towards numbers and letters, and general motivation. Each item was scored on a three-point smiley face scale or small to large circles for questions pertaining to emotions or attitudes, respectively. Higher scores indicated more positive attitudes while lower scores indicated negative attitudes (i.e. happy to play numbers games would score higher and indicate positive outcome, see Figures A1-3 in appendix).

***Executive Function. Corsi Blocks Task (Corsi, 1972, as used by Blakey et al., 2020).*** This is a measure of children's visuospatial short-term memory. Nine 2x2cm wooden blocks were attached to a white A4 piece of cardboard in a random array. The researcher tapped two blocks in a pre-set random order and the child was instructed to tap the same two blocks. For each numerosity, the child completed 3 trials. If 2 or more sequences were correct, the child progressed onto the next numerosity. Each trial was coded for whether the blocks tapped were correct or incorrect and whether the child had tapped the blocks in the correct sequential order or not. The variable used for analysis was the overall number of correct trials, regardless of sequential order. The test-retest reliability of this task has been shown to be very good (ICC: .90; Alloway & Passolunghi, 2011) and it has been successfully used in number of studies on preschools (e.g. Blakey et al. 2020). ***Mr Ant (Early years toolbox; Howard & Melhuish, 2017)*** is a visuo-spatial memory task presented on a tablet in which the child is asked to remember the location of colourful stickers on a cartoon image of an ant. In each trial, the stickers are presented for 5s, followed by a blank screen for 4s. A blank ant reappears and the child is asked to indicate where the stickers had previously been by tapping. There are three trials in each block, with the child progressing to the next block if he or she is correct on at least one trial. The number of stickers that appear start at 1 in block 1 and increase by 1 in each block. A score was calculated, with one point given for each trial with 2 or 3 items correct. On trials with 1 item correct, 0.33 points were added to the score (following Howard & Melhuish, 2017). This task was normed on Australian preschools (Howard & Melhuish, 2017), and has been used in British preschoolers (Coolen et al., 2021). It is sensitive to change over a period of 5 months. ***Rabbits & Boats (Howard & Melhuish, 2017)*** is a tablet-based shifting task, based on a traditional card sort task (Zalazo, 2006). Across three blocks, the child must sort cards first according to colour (red/blue), then according to shape (rabbit/boat), and finally switching the rule depending on whether or not there is a black border. Each block contains 6 trials and the child must get at least 5 trials correct on block 2 in order to progress to block 3. A switch accuracy score, calculated as the sum of the total scores for blocks 2 and 3, was used for analysis (following Howard & Melhuish, 2017). ***Fish-Shark Go/No-Go (Howard & Melhuish, 2017)*** is a tablet-based task of inhibitory control. Either fish or sharks moved across the screen, and the child was instructed to tap the fish (go trials) and not tap the sharks (no-go trials). There were 3 blocks of 25 trials, each consisting of 20 go trials and 5 no-go. In the data cleaning process, data were removed for trials with (a) a response time of less than 300ms, (b) non-responsiveness (go accuracy of <20% and no-go accuracy >80%) (c) indiscriminate responsiveness (go accuracy of >80% and no-go accuracy of <20%). After these trials had been removed, proportional go and no-go accuracy scores were multiplied to create an overall impulse control score, which was used for analysis (following Howard & Melhuish, 2017). Scores from these four executive measures were entered into an exploratory factor analysis.

### Intervention Co-development.

In order to create activities that were feasible and acceptable to practitioners, whilst containing appropriate mathematical and executive function content, several stages of development and refinement took place. At ***Stage 1 (initial development with cognitive scientists and educators)***, an initial set of activities were drafted, using inspiration from a self-regulation intervention PRSIST (Howard et al., 2018) and a primary school mathematics intervention (Hawes et al., 2021). Activities were refined and modified with the advice of advisory board members, to ensure that they contained sufficient EF challenge and numerical content in line with early years foundation stage guidelines (Department for Education, 2020). For example, following the advice of an early years practitioner on the advisory board, the mathematical content of some activities was scaled back at this stage (e.g., using numerals 1-5 instead of 1-10), whilst retaining the executive challenge. ***Stage 2 (piloting with 4 Early Years settings).*** Piloting took place with four early years settings, chosen to represent a range of preschool characteristics (SES, size, type, location). These settings were each given all twenty-five activities, in smaller weekly packs of six or seven activities and asked to provide feedback on three activities per week both in writing using a feedback sheet and verbally in weekly feedback sessions. Activities were refined based on this feedback (e.g. removing calculation element from certain activities). ***Stage 3 (continued refinement with 8 intervention settings).*** During the main trial, written feedback was collected from each of the eight intervention settings for each activity carried out over the 12-week duration of the intervention using a feedback poster which collected data on the number of children involved, a 3-point smiley face rating for activity success and additional comments. This quantitative information was supplemented by qualitative interview questions about the most and least successful activities and practitioners' perceptions of why some activities were more successful than others (see table A1, in the Appendix). Activities that had either a low popularity rating, low frequency of selection or were highlighted as unsuccessful at interview were reviewed. Based on qualitative feedback from interviews and comments on the feedback sheet, some additional changes were made to activities. Most changes were related to ensuring activities were accessible for different ability levels (e.g., replacing language element from *what number am I?* with manipulatives for counting), non-repetitive (e.g., combining three relational rods activities into a single activity), appealing for children (e.g., using 3d straws as manipulatives in place of strips of paper), and clear for practitioners (e.g., changing the name of *addition/subtraction ball* to *throwing with numbers*). A general theme that emerged from interviews was the suitability of activities for children with special educational needs, English as an additional language or lower ability. ***Stage 4 (continued refinement with 7 BAU settings).*** In response to this feedback, a "top tip" box was added to each instruction card, containing advice for adjusting the activity. These changes will be evaluated during re-delivery to the control group settings, which is currently ongoing and will be summarised in a manuscript for publication (O'Connor et al., in preparation).

### Main Feasibility Trial Intervention Protocol.

There were weekly 30-minute face-to-face professional development sessions for the first four weeks of the programme. These sessions introduced the 25 activity cards, supported practitioners' explicit understanding of how early mathematics and executive functions co-develop, and explained how executive functions can be embedded into a range of early mathematics learning activities (number and counting activities, but also space-focused activities and ordering activities). The research team asked all staff in a setting to attend these sessions if possible. Nursery staff were also asked to practice a minimum of three activities per week with children due to enter Reception the following September. There were two additional requests. First, weekly activities should be chosen to target all three key areas of mathematics represented in the activity pack (numbers and counting, patterns and ordering, space and shapes). Second, after children were familiar with activities, practitioners were asked to increase the executive challenge of each activity. One representative per setting was contacted in the eighth and twelfth weeks by the delivery team, to provide support, an opportunity for self-reflection and to check fidelity of delivery.

### Practitioner-level Assessments.

***Feedback posters.*** Feedback sheets (week 1-3) and a poster (week 4-12) were used as a simple way of measuring adherence. For each of the 12 intervention weeks, practitioners were asked to record the activities completed, how many children attended, activity success using a three-point smiley face scale, whether additional challenge was added, and additional comments. Space was provided for three activities per week, with an extra sheet provided for any additional activities.

***Observations.*** One activity observation was carried out at each preschool, with duration of activities ranging from 10 to 25 minutes. Educators were advised to choose an activity prior to the observation. No directive guidelines were given on activity type or group size, but we observed a spread of activities implemented across settings. As the intervention activities were designed to be adaptable to the needs of a specific setting, a strict, categorical observation scheme was not deemed to be appropriate. The aim of the observation scheme was instead to measure adherence to the goals, quality of delivery and ability to adapt activities appropriately for the children present.

***Practitioner Evaluation forms.*** After 4 weeks of PD sessions, practitioners were asked to complete an evaluation form containing questions related to the success of PD training and of activities. Each form contained 7 statements related to PD training and 7 statements related to the activities to rate from 1 (strongly disagree) to 5 (strongly agree). Space was also provided for qualitative insights. 23 forms were returned across all 8 intervention settings (Mean forms per setting = 2.88, range = 1-6).

***Interviews.*** Interviews were conducted by the researcher with one practitioner at each preschool at the end of week 12 of the intervention, each lasting approximately 30 minutes. As interviews often took place within the nursery room, it was not appropriate to collect audio data. Instead, detailed coding sheets were used, allowing the researcher to code key variables and write notes as the interview was conducted.

## Data Treatment and Data Analysis Plan.

We pre-registered our primary and secondary outcome variables on Open Science Framework (https://osf.io/8y5u6/ ). **Our primary outcome measures were acceptability, feasibility, adherence and fidelity of the intervention.** In brief, we committed to collecting **acceptabilit**y ratings from any settings exiting the study and endline acceptability questionnaires from those who took part in the intervention (including qualitative open-ended questions and quantitative ratings on practitioner training sessions and intervention activities). For **feasibility**, we committed to collecting rates of enrolment and retention, and logistical problems based on logs kept by practitioners. We also aimed to collect qualitative information on feasibility (from open-ended questions from practitioner interviews). For **adherence and fidelity**, we aimed to require attendance to practitioner training of 75% training sessions, by 75% of staff at each setting; and 75% adherence to delivering the intervention activities (a minimum 3 per week). We also committed to report on qualitative observations of intervention fidelity in delivery at the end of the intervention period (weeks 11 – 12 in each setting) and to computing implementation quality indicators as recommended by Dowling and Barry (2020). We proposed that these primary outcome data will be descriptive, not inferential.

**Our secondary planned outcome measures were indices of preliminary intervention efficacy** on children's early mathematics and EF from pre- intervention to post-intervention. We planned to use two-way mixed ANOVA to analyze the intervention effects on the secondary outcome variables. We aimed to employ the traditional $p < .05$ frequentist convention. In case of null intervention effects, we aimed to follow Bayesian approaches, again following conventions for confidence in the effects detected (in progress). However, after data collection, we found data distributions violating assumptions of parametric statistics (normality), baseline differences despite pseudo-randomisation to the treatment arms and missing data; this required approaches that deviated from the pre-registered analyses. An intention-to-treat approach was used, therefore, including all children who had been exposed to the intervention into the analyses. Multi-Level Linear Modelling (MLM) is deemed appropriate for $N > 100$ provided variables are normally distributed (Snijders & Bosker, 2012) and was therefore employed whenever possible. Group (BAU, Intervention) Quade non-parametric Analyses-of-Covariance (ANCOVAs) were carried out, controlling for differences at Time 1, when variables did not meet assumptions of normality and transformations did not rectify violations. For numeracy, the cumulative early years mathematics measure (Early Years Toolbox Numbers, "EYTN" henceforth) was analyzed first, followed by individual component mathematical skills. For Executive Functions, data reduction (exploratory factor analysis) was employed to produce latent EF factor score and reported first, followed by individual executive functions measures. In addition, analyses modelled the effects of Early Years Pupil Premium (EYPP) eligibility (EYPP; Yes, No, Unknown), to investigate potential differential effects of the Intervention for disadvantaged children. Our further exploratory goals were to develop a new child-report measure of attitudes towards mathematics and learning, based on related constructs that have been measured in older children, test the internal consistency of the new scale, and test its correlation with early years mathematics and executive function measures collected as part of the RCT. Here we opted not to report estimates of effect size and confidence intervals, given the small sample size of this feasibility RCT.

# Results

## Primary Outcome Variables.

### Acceptability.

An initial index of acceptability of The ONE Programme came from the recruitment phase. Fifty-seven settings were initially approached by the research team to introduce the background to the study. Forty-eight settings expressed an interest in receiving full study information. Out of these, 22 settings volunteered in full to take part in the study. Two settings withdrew prior to scheduling pre-intervention assessments and one setting withdrew while pre-intervention assessments were ongoing. In these three cases, staff time pressure and COVID were reported as the primary reason for withdrawing.

Further indices of acceptability were obtained from interviews at the end of intervention delivery. As a whole, multiple benefits were reported for practitioners and children (see Table 2). At the same time, staffing issues and adaptability to children with special educational needs (SEN) or English as an additional language (EAL) were reported as the most frequent barrier to the programme.

**Table 2. Themes emerging from interviews**

| Theme | Num. Settings | Examples (anonymised by setting) |
|---|---|---|
| *Benefits of the programme* | | |
| Useful activities | 5 | *[the activities] helped the most when I needed structure and calm in the classroom*<br>*The activities were all fun for the children, and it was interesting to see how much variability there was in the children's performance* |
| Staff attitudes and behaviour | 4 | *The ONE helps practitioners to understand where the children's skills are – we are surprised sometimes!*<br>*We were already mathematics focused, but now we have more ideas and more purpose.* |
| Staff knowledge development and PD | 3 | *Mathematics has always been a priority in our routine, so there was not a major change. We are thinking more about executive functions. We know that just a few tweaks can make an activity more challenging.*<br>*The training was really interesting and eye-opening – I learnt lots of new things and it has made me observe the children in a new light.* |
| Children's skill development | 3 | *The children's confidence improved as well as their mathematics skills*<br>*Challenge became less scary*<br>*It's all to do with the skills they will need when they start school – remembering, counting, sitting still, listening, teaching each other, communication, subitising, shapes, words, preposition words* |
| Suitable for children with SEN | 1 | *[SEN child] enjoyed numbers and sequences, number line worked well for her to practice inhibition and gave her a visual aid* |
| Other benefits for staff | 1 | *This was a great extra learning opportunity as part of my level 3 EY diploma* |
| *Barriers to taking part* | | |

| | | |
|---|---|---|
| Staffing issues | 6 | *The timing was hard because of annual leave and sickness – but it was helpful to have activities that could group children together and adapt to their abilities in one go* |
| Child ability | 4 | *Some activities were better than others. The content was not always appropriate for EAL and SEN children.* |
| Reporting on activities | 3 | *I didn't love reporting – it was difficult to do right after activity and difficult to remember to do it afterwards*<br>*We struggled with filling in the paperwork. Maybe an electronic version would work better. We already take photos to track the children's progress* |
| Covid-19 pressures | 2 | *The children's behaviour has been more challenging since the covid-19 pandemic* |
| Planning and time | 2 | *Training sessions should be longer to allow time for planning. A basic problem was that there was no time to communicate and plan within our team across different rooms.*<br>*There are constant changes in our routine, so there's no point trying to plan* |
| Preschool ethos | 1 | *The main barrier was the ethos of the preschool. We have no time to plan activities and the children are supposed to spend all day doing free play – there is no time for organised adult-led activities* |
| Staff capacity and motivation | 1 | *We have had staff difficulties, with less commitment and buy in from some newer colleagues*<br>*It is difficult to include agency staff in the intervention and core staff have limited time* |

Finally, acceptability indices came from practitioners' feedback on the play-based activities. The 25 activities were each rated with a score out of 3. The average mean rating per activity was high, 2.7 (SD = 0.26, range = 2.3 – 3.0).

## Adherence and Quality of Implementation

**Quantitative measures**

***Adherence:*** The percentage of activities completed over 12 weeks out of a maximum of three per week was calculated, as reported by practitioners in the posters and feedback sheets. This revealed that 5 settings met the >75% adherence to delivering 3 intervention activities per week, whilst a further 3 settings did not meet this criterion (see Table 3). All settings attended all four PD sessions, thus meeting the >75% sessions criterion. The final criterion of >75% staff attendance at PD proved difficult to measure due to the variable nature of roles, responsibilities and contract types for preschool staff. Nonetheless, at least two members of staff were present at each PD session at each preschool, so all settings adhered on the grounds of this criterion.

**Table 3. Adherence to the intervention across settings**

| Setting | Activities completed (out of a possible 3 per week) | Additional activities | Number of PD sessions attended | Average (SD) number of staff at each session |
|---|---|---|---|---|
| A | 100% | 3 | 100% | 3.00(0) |
| B | 100% | 9 | 100% | 2.00(0) |
| C | 100% | 5 | 100% | 2.50(.500) |
| D | 97.2% | 4 | 100% | 5.00 (0) |
| E | 86.1% | 4 | 100% | 5.00(0) |
| F | 72.2% | 0 | 100% | 5.75(.433) |
| G | 63.9% | 0 | 100% | 5.75(.829) |
| H | 25.0% | 0 | 100% | 2.00(0) |

***Implementation Quality Indicators:*** Following a similar method to Dowling and Barry (2020), appropriate indicators were selected from the adherence posters, observations, teacher evaluations and interviews to represent the dimensions of implementation: Dosage, Fidelity, Quality of Delivery and Participant Responsiveness.

***Dosage:*** A single indicator of practitioner-reported activity completion was used. The percentage of activities completed over 12 weeks out of a maximum of three per week was calculated, as reported by practitioners in the posters and feedback sheets. ***Fidelity:*** Fidelity was computed using 2 indicators from the observation sessions. Prior to the observation, the researcher listed five key features of the activity. Each of these key points were coded (yes/no) depending on whether they had been observed in the activity, and a final percentage score was calculated based on the number of points achieved out of a possible five. A second indicator of fidelity was calculated using four items on a Likert scale related to adherence to the activity card (e.g., "*The activity contained appropriately challenging executive challenge, as outlined on the activity card."*). These 4 items were averaged and a percentage score (out of 7) was calculated. ***Quality of Delivery:*** Quality of delivery was computed using 2 indicators. During observations, any adaptations from the original activity were noted and coded for whether the change had a positive or negative effect on the activity. A mean of 3.63 (SD = .992) adaptations were carried out per setting. As an implementation quality indicator, a percentage score of positive adaptations was calculated for each preschool. A second observation-based indicator of fidelity was calculated using eight relevant items on a Likert scale (e.g., "*The practitioner gave clear instructions and/or demonstrations to ensure that the children understood the game well."*). These 8 items were averaged and a percentage score (out of 7) was calculated. ***Participant Responsiveness:*** Participant responsiveness was calculated using 3 indicators. During the observations, a conditions of delivery score was calculated using 5 items on a Likert scale (e.g., "*The activity was too difficult for the children present*."). These 5 items were averaged and a percentage score (out of 7) was calculated. Next, a practitioner reported measure of activity success was calculated based on responses on the evaluation form (e.g., *"The activities were fun for the children."*). Items were scored from 1 (strongly disagree) to 5 (strongly agree). This was then collapsed into a three-point scale to avoid individual differences in scoring style affecting the final score and a percentage score was calculated and averaged across settings. Finally, during interviews practitioners were asked about the impact of the intervention upon the children (Positive/Neutral/Negative). These answers were converted to a percentage score out of 3 (1/3 = Negative, 2/3 = Neutral, 3/3 = Positive).

**Analysis**

Following the method outlined by Dowling and Barry (2020), an average of the four dimension scores (dosage, fidelity, quality of delivery and participant responsiveness) was calculated, proving a total score for implementation quality of The ONE. These scores were then grouped using the mean and ±.5 SD as cut-off points. These cut off points were deemed more appropriate than mean and ±1 SD as, the cut-off point for the upper bin was higher than 100%. This resulted in three groups: high (N = 4), moderately high (N=2), and low (N=2) (Table 4). No preschools fell into a "moderately-low" bin.

**Table 4. Mean and range scores for each implementation quality dimension, split by overall group**

| Dimension | High Group (N = 4) | Moderately High Group (N = 2) | Low Group (N = 2) | Total (N =8) |
|---|---|---|---|---|
| Total Dosage Score | 95.8% (86.1% - 100%) | 68.1% (63.9% - 72.2%) | 62.5% (25.0% - 100%) | 80.6% (25.0% - 100%) |
| Total Fidelity Score | 94.4% (81.1% - 100%) | 97.3% (96.4% - 98.2%) | 27.5% (17.1% - 37.9%) | 78.4% (17.1% - 100%) |
| Total Quality of Delivery Score of The ONE | 97.5% (95.5% - 99.1%) | 96.5% (93.8% - 99.1%) | 51.8% (35.7% - 67.9%) | 85.8% (35.7% - 99.1%) |
| Total Participant Responsiveness Score | 99.3% (98.2% - 100%) | 96.5% (93.8% - 99.1%) | 48.7% (29.0% - 68.3%) | 87.2% (67.4% - 100%) |

| | | | | |
|---|---|---|---|---|
| Total Implementation Quality Score of The ONE | 96.5% (94.2% - 98.4%) | 85.3% (84.6% - 85.9%) | 48.7% (29.0% - 68.3%) | 83.0% (39.0% - 98.4%) |

**Qualitative data**

Interviews and evaluation forms were written up and coded for emerging themes (see Figure 2). As Figure 2 shows, themes clustered around benefits of the programme, barriers to taking part, more successful and less successful activities, and suggestions for the future.

***High implementation quality (N = 4).*** Four settings showed high implementation quality. Each of these settings also reported doing extra activities in addition to the suggested 3 activities per week (see table 4). These settings reported numerous benefits of the programme, both for staff (e.g. *"We were already maths focused, but now we have more ideas."*) and for the children (e.g. "*The children's confidence improved as well as their mathematics skills"*). At interview, each of these settings reported changes in the preschool staff (e.g. *"[the activities] helped practitioners to think more about challenge in activities and provided inspiration", "The ONE helps practitioners to understand where the children's skills are – we are surprised sometimes"*). The main barriers to taking part mentioned were staff shortages and remembering to report activities.

***Moderately high implementation quality (N = 2).*** Two preschools showed moderately high implementation quality. Both these settings showed low adherence but performed well on observation measures. During the interviews, both interviewees cited planning and issues with staff co-ordination as being barriers to fully engaging with the project, "*Training sessions should be longer to allow time for planning. A basic problem was that there was no time to communicate and plan within our team across different rooms."* In the evaluation forms and interviews, activity suitability for lower ability children was a common theme at both settings, e.g. *"The content was not always appropriate for EAL and SEN children".*

***Low implementation quality (N = 2).*** Two settings showed low implementation quality. Interviews revealed complex issues at both preschools, including a high turnover of staff, some unwilling staff members, a lack of compatibility with the ethos of the preschool, issues with planning and a lack of time. Educators at these settings generally found the training informative, making comments such as *"Very good and informative*" and *"[I would like] more time on the training"*, but struggled to find time to complete activities with the children *"We didn't have enough time to go through all the activities, however they seemed to be very good and informative".* Evaluation forms also revealed that some staff found it difficult to understand and explain concepts, e.g. *"[a problem that I experienced was] understanding how to explain".*

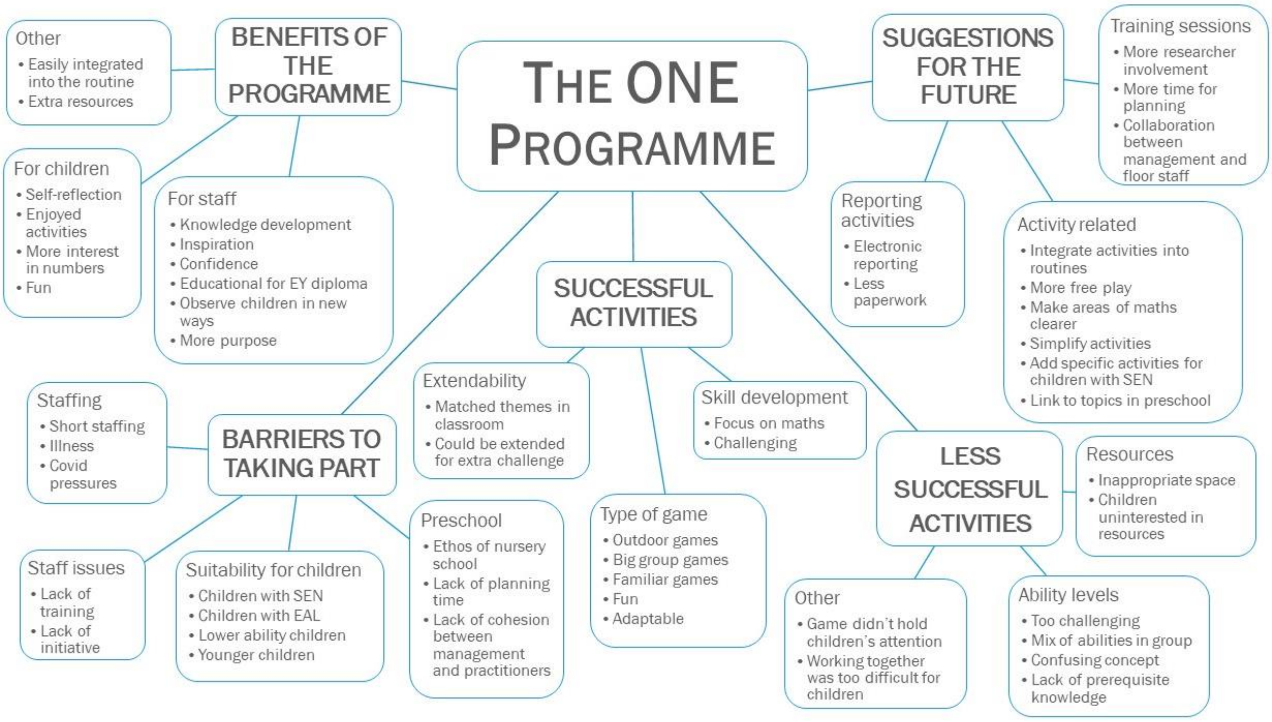



*Figure 2. Themes emerging from interviews with practitioners*

## Measures of Practitioner Change: Mind Maps

Mind maps were used as a simple measure of subject knowledge development pre- and post-PD. Practitioners were given blank mind maps containing a key term (*Early numeracy skills* or *Executive functions*) and were asked to populate the mind map with any relevant terms that came to mind (see Figure 3). These mind maps were completed as an exercise at the start of week 1 of the PD sessions, and again during week 4 (the final week) of PD. Due to the variable nature of preschool settings, these were completed individually in some settings (N = 4) and in a group at some settings (N = 3). At one setting, the pre-PD mind map was completed as a group and the post-PD mind maps were completed individually.

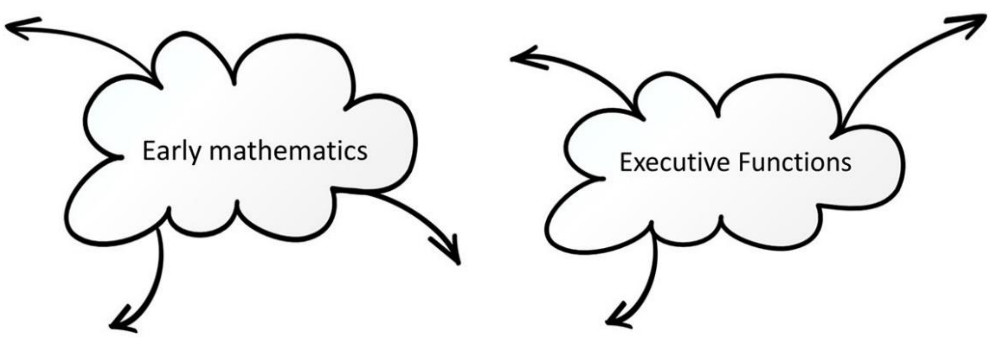



*Figure 3. Example mind maps for two key constructs*

**Quantitative**

The number and accuracy of spokes were used as quantitative measures of mind map content. Spokes on each mind map were transcribed and coded as either accurate or inaccurate. Spokes were coded as accurate if they described the theme or were clear examples of a behaviour relevant to theme (e.g. *avoiding distractions* for executive functions, or *counting games* for early mathematics). Inaccurate spokes were unclear, irrelevant or non-specific (e.g. *play* for executive functions, or *rhyming* for early mathematics).

*Results*

As the sample size is small (N=8), all statistics reported will be descriptive rather than inferential.

Overall, more spokes were produced for early mathematics than executive functions both before PD (Mean $_{EF}$ = 2.33, SD = 2.39; Mean $_{Maths}$ = 6.25, SD = 2.77) and after PD (Mean $_{EF}$ = 3.88, SD = 1.73; Mean $_{Maths}$ = 6.00, SD = 3.48). For executive functions, the number of spokes increased from T1 to T2, as well as the proportion of accurate spokes (from 34% to 58% see Figure 4). For early mathematics, there was an increase in the accuracy (from 73% to 89%) but minimal in the total number of spokes (see Figure 4).

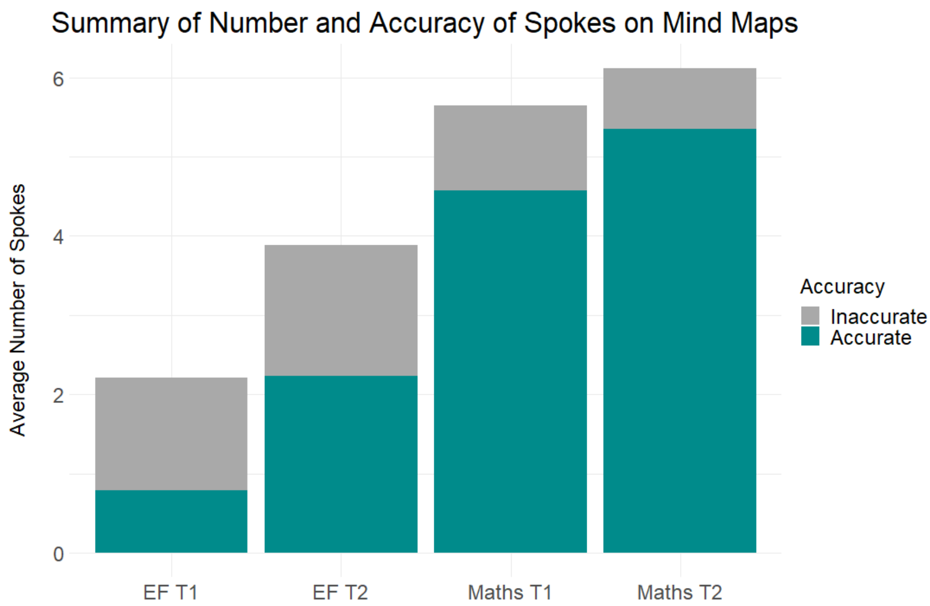


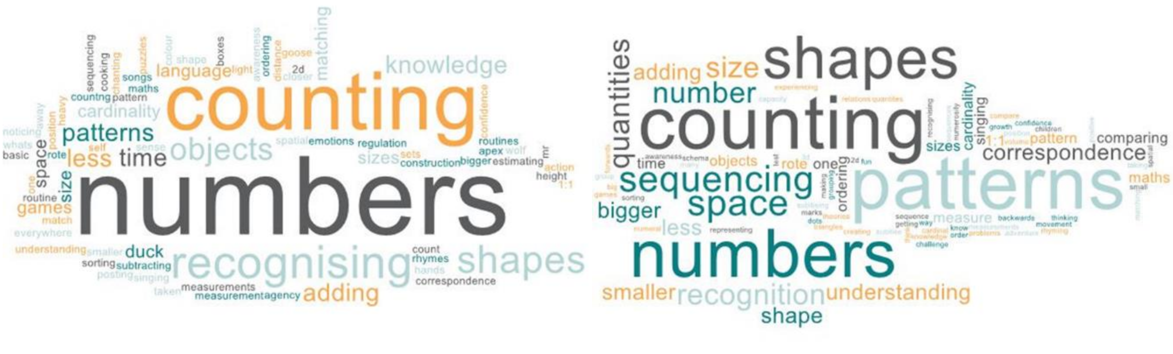

*Figure 4. Changes in the number and accuracy of mind map spokes for math and EF at T1 (pre-PD) and T2 (post-PD)*

**Qualitative**

Word clouds were generated as a way of visualising the qualitative differences in mind map responses between T1 and T2 (see Figures 5-8). For EFs, the T1 word cloud is largely dominated by *memory* and general words related to cognition, such as *brain, processes* and *think.* By T2, although *memory* and *thinking* are still dominant, there is more of an even spread of words related to other areas of EF and to behaviour regulation more generally: *focus, inhibition, problem (solving), rules, regulation* and *turn taking.*

Although there were few quantitative differences between the mind maps at T1 and T2 for mathematics, the words clouds reveal qualitative differences between the two time points. Despite the fact that there is a good spread of relevant terms at T1, there is a clear dominance of *counting* and *number/numbers*, followed by *shapes.* In the T2 word cloud, there are now four roughly equally dominant terms: *counting, number, shapes* and *patterns*, which roughly map onto the three main areas of maths covered in the intervention. Following these are several moderately high frequency terms, which cover a greater spread of mathematical areas, for example: *space, sequencing, (number) recognition, (1:1) correspondence, bigger/smaller* and *size.*


*Figures 5 and 6. Word cloud of practitioner responses to the prompt of "executive functions" prior to PD (on the left) and post PD (on the right).*




*Figures 7 and 8. Word cloud of practitioner responses to the prompt of "early mathematical skills" prior to PD (on the left) and post PD (on the right).*

## Secondary Outcome Variables.

### Intervention Efficacy.

***Summary.*** The intervention group achieved higher scores at Time 2 than the BAU group on the Early Years Toolbox Numbers (EYTN) scale, a combined numeracy measure, and higher scores than the

BAU group on Corsi Blocks, a short-term memory scale, after controlling for Time 1 differences. In addition, modelling EYPP eligibility highlighted interactions of Intervention with EYPP. Children with EYPP eligibility in the Intervention group scored higher than children with EYPP eligibility in the BAU group on EYTN, on number comparison skills, and on spatial skills (BAS Pattern Construction tscores). Children with EYPP eligibility in the Intervention group scored higher on the EF latent score and on two short term memory measures (Corsi Blocks and Mr Ant) compared to BAU children with EYPP. On Counting high, Intervention children scored higher than BAU children, but only in the group of children who were not eligible for EYPP. This summary is supported by the statistics below.

***Overall Intervention Effects*****.** *Effects on Numeracy.* Overall Numeracy Scores (Early Years Toolbox Numbers, EYTN score) and British Ability Scales Pattern Construction (BAS-PC) tscores were normally distributed, allowing for modelling via MLM. For EYTN score, there were main effects of Time ($p < .001$, with scores improving from Time 1 to Time 2) and of Intervention Arm, $p < .019$, driven by higher overall numeracy scores for children in the Intervention Arm (high adherence) compared to children in the Intervention Arm (low adherence, $p = .008$), and compared to children in the BAU arm ($p = .041$), having controlled for Time 1 scores (please see Figure 8). For BAS-PC tscores, there was a main effect of Intervention, $p = .004$, driven by higher scores for children in the Intervention Arm (high adherence) compared to children in the Intervention Arm (low adherence), $p = .001$, and a trend to higher scores compared to the BAU arm ($p = .102$). Other numeracy variables were not normally distributed and were therefore analyzed with non-parametric Quade ANCOVAs, but these analyses did not reveal statistically significant effects of Intervention. *Effects on Executive Functions.* Exploratory factor analysis on the four EF task scores yielded a single factor with an Eigen value greater than 1. The EF latent factor was normally distributed, and therefore MLM modelling was appropriate. There were no statistically significant effects on this variable. All other variables were not normally distributed, so Quade ANCOVAs were used. For Corsi Blocks (working memory), there was a main effect of Intervention effect, $p = .044$, after controlling for Time 1 scores, driven by a bigger improvement for the Intervention Group compared to the BAU Group. There were no other statistically significant main or interaction effects with Intervention. Please see Table A6 and Figures A4-A22 in the Appendix for all descriptive statistics.

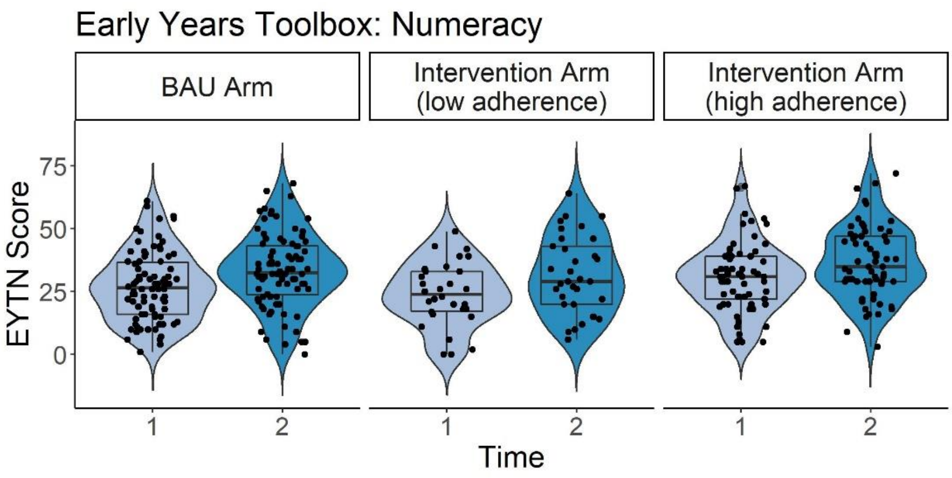



**Figure 8.** EYTN scores plotted by intervention group, and pre- post intervention scores. Individual dots represent individual children.

***Intervention Effects modelling EYPP eligibility.*** *Effects on Numeracy.* There were main effects of EYPP, Intervention Arm and Time for overall numeracy (EYTN score) and for spatial skills. For EYTN score, there was a main effect of Intervention Arm, p = .046 (with higher score for Intervention children), a main effect of EYPP eligibility (with lower scores for EYPP eligible children compared to EYPP not eligible children and children whose status was unknown), p <.001, a main effect of Time, p = <.001 (with higher scores at Time 2 compared to Time 1), and an Intervention * EYPP interaction effect, p = .044. The interaction effect was driven by the following differences: for children with EYPP eligibility, EYTN scores were higher in the Intervention group than in the BAU group (p =.005). In addition, children with EYPP eligibility scored less well than children without EYPP in the BAU group (p <.001), but this difference was reduced for children in the Intervention group (p = .012) (See Figure 9).

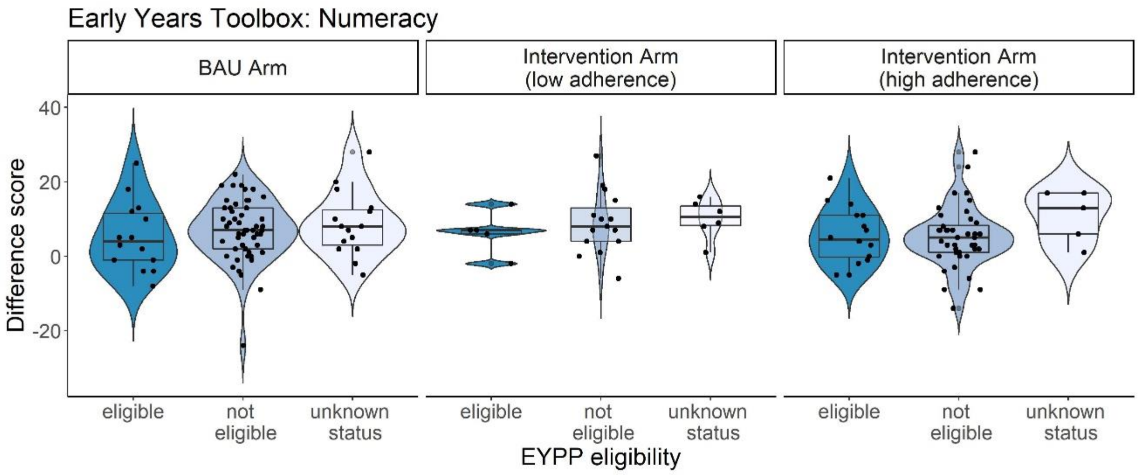



**Figure 9.** EYTN difference scores, plotted by intervention group and EYPP eligibility.

For spatial skills (BAS3-PC), there was a main effect of EYPP eligibility, $p < .001$ (with lower scores for EYPP eligible children) and an Intervention * EYPP eligibility, $p = .024$. The interaction effect was driven by the following: children with EYPP eligibility had poorer spatial skills than children without EYPP eligibility in the BAU group, $p <.001$, but not in the intervention group, $p =.218$. In addition, children with EYPP in the Intervention arm had better spatial skills than children with EYPP in the control group, $p =.010$. All other numeracy variables were not normally distributed, and therefore MLM was not appropriate. Effects were assessed using non-parametric Quade ANCOVAs. For Digit Comparison (proportion correct), there was a main effect of Group, $p = .020$, after controlling for Time 1 scores, again with poorer number comparison skills for children with EYPP than those without in the control group, $p = .003$, but not in the Intervention Group, $p =.111$. .
*Effects on Executive Functions.* There was a main effect of EYPP eligibility on EF latent scores, $p < .001$, and a trend towards an Intervention * EYPP interaction effect, $p = .063$. This was driven by lower scores for children with EYPP status than those without in the BAU group, $p < .001$, but a smaller difference for children with EYPP status compared to those without in the intervention group, $p = .025$, and a trend towards higher EF latent scores for EYPP eligible children in the Intervention arm compared to children with EYPP in the BAU arm, $p = .069$. The individual EF variables were not normally distributed. For Corsi Block Scores, there was a main effect of Intervention Arm, $p = .044$, controlling for Time 1 scores (Quade ANCOVA), driven by higher scores in the intervention compared to BAU group. For Mr Ant Scores, the other working memory index in our EF measures, there was a main effect of Group, $p = .006$ (Quade ANCOVA) after controlling for Time 1 scores, driven by lower scores for children with EYPP compared to those without in the BAU group, $p = .003$, but not the Intervention Group, $p =.419$. Please also Table A7 and Figures A4 – A22 in the Appendix for descriptive statistics for all variables.

## Additional measure development.

*Child Attitudes Questionnaire – Preschool (CAQ-P): Design and Outcomes.* This questionnaire was designed to target the existing gap in the literature on children’s attitudes towards maths in the early years. While literature supports the co-development of academic emotions and motivation alongside educational outcomes in school (Dietrich et al., 2022; Obersteiner, 2018), research is limited on how this happens before the start of formal schooling. We aimed to design a questionnaire which adequately measured emotions towards maths, attitudes towards maths and children’s general motivation (Dowker et al., 2019). A key goal was to design questions to be construct-specific while still adhering to the vocabulary knowledge of young children (Ganley and McGraw, 2016). For example, one might use the word “anxious” when testing adolescents, but one cannot in young children and we therefore use the word “worried” to describe a negative feeling associated with negative emotional arousal when doing maths activities. Another example is using “numbers” rather than “maths”. The questionnaire consisted of a total of 23 items with nine items being maths specific. Experimenters reported on whether children understood the words used, the scale used to answer and the constructs they were asked about. Finally, we examined whether CAQ-P scores were related to numeracy and EF performance.

The final N consisted of 165 children after removing those who had only completed portions of the questionnaire or were absent at the time of administration. 76% of children understood what the word “worried” means, 91% understood what the word “happy” means, 90% understood what the word “numbers” means and 83% understood what “letters” means. Furthermore, 80% understood the worried faces scale and 85% understood the happy faces scale. We preliminarily saw strong internal consistency for items on the Emotions subscale (12 items) at a Cronback’s Alpha score of .755 but not for the Attitudes (6 items) or Motivation (5 items) subscales, although all three encompassed correlates of numeracy and EF. Further, we found that children were able to respond differently at both the domain specific (emotions vs attitudes vs general motivation) and construct specific (number emotions vs word emotions and number attitudes vs word attitudes), see Figure 10 for correlations. Finally, the most important question was to understand whether children could report on their feelings about numbers this early on. Checks in place to ascertain children’s understanding and cross construct comparison gave confidence that children understood the number specific questions. Importantly, we found that children’s emotions towards numbers correlated with their EYTN scores at .380, $p<.001$. Further, children’s emotions towards numbers also correlated with their impulse control and working memory (WM) performance, at .243, $p=.005$ and .183 with $p=.028$ respectively. This is interesting, because impulse control and WM performance in later childhood and adolescence influence the relationship between maths anxiety and maths performance (Dakin, Gattas et al., 2022).

These results suggest that children’s emotions and attitudes towards numbers and letters are developing at the earliest stages of schooling and may be foundational to their educational and specifically, numeracy development. It would prove beneficial to include emotions, attitudes and motivation in future maths interventions to address how we can best equip children with foundations upon which they start to build their maths knowledge.

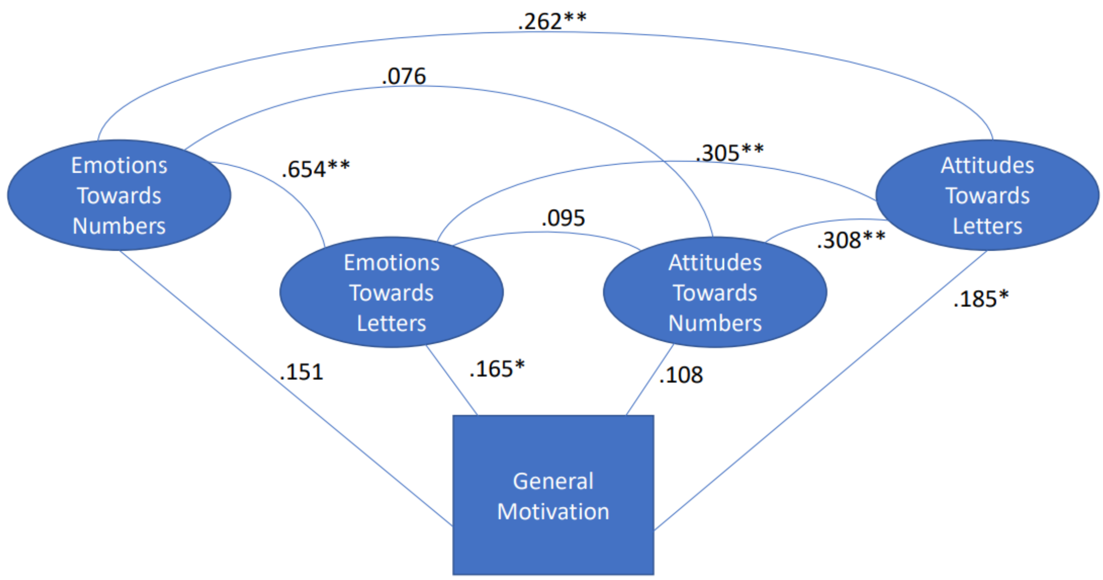



Figure 10. Correlations between subscales of the Child Attitudes Questionnaire – Preschool (CAQ-P).

# Discussion

## Outcomes and implications.

The current study aimed to bring together cognitive scientists and educators to co-develop and progressively refine an evidence-based integrated Executive Functions and Mathematics intervention composed of professional development and play-based activities. This small scale feasibility RCT showed evidence of good acceptability and feasibility according to multiple metrics. All preschools had appropriate attendance at practitioner development sessions and most showed acceptable adherence in carrying out activities. Importantly, initial measures of knowledge development indicated both quantitative and qualitative gains in understanding of key intervention terms. Observations demonstrated that at most preschools, staff were able to put this knowledge into practice to lead high quality activities that met the key aims of the intervention. Qualitative data from interviews indicated that overall the intervention was regarded to be beneficial for both staff and children. These outcomes for early years educators have positive implications, given previous calls by early years educators for professional development in the area of early numeracy (e.g., von Spreckelsen et al., 2019).

Our secondary target was an evaluation of efficacy of The ONE in improving early numeracy outcomes for children in the intervention. Previous research points to concurrent and longitudinal correlations between early numeracy and executive functions (Coolen et al., 2021), but interventions that have focused on executive functions in isolation have tended not to result in improvements in early numeracy. We have hypothesised that interventions integrating executive challenge with mathematical content have the potential to improve early numeracy most effectively (Scerif et al., under review). Indeed, The ONE resulted in greater differential change on a combined numeracy measure (EYTN) for children in the intervention arm compared to those in the BAU arm. Children in the intervention arm also outperformed the BAU group on a memory scale (Corsi Blocks). In addition, children who were more economically disadvantaged (as indexed by EYPP eligibility) but were in the Intervention group scored higher than children with EYPP eligibility in the BAU group on overall numeracy EYTN and on some foundational early numeracy skills (number comparison skills and spatial skills). Children with EYPP eligibility in the Intervention group scored higher on the EF latent score and on two short term memory measures (Corsi Blocks and Mr Ant) compared to BAU children with EYPP. As a whole these preliminary efficacy data suggest that The ONE had small effects on the intervention group as a whole, but that intervention benefits were larger for disadvantaged children, who needed these improvements most.

Finally, a new measurement tool was developed to assess preschool children's attitudes towards early numeracy, and fill a gap in the existing literature at a time when attitudes towards learning may be critical (Dowker et al., 2019). Emotions towards numbers correlated with early numeracy scores, impulse control and working memory scores, suggesting that children's emotions and attitudes can be meaningfully measured as early as in preschool.

## Future Directions.

Together with these positive outcomes, it is important to highlight considerations for future scaling up of The ONE, particularly with regards to facilitating adherence and refinement of activities that were reported as less age-appropriate than others. Future refinement steps should focus: 1) on reducing staff time pressures as much as possible, 2) easing planning so that The ONE activities can be even more easily integrated in setting planning, 3) easing reporting on adherence, and 4) revisiting specific activities to incorporate feedback from this phase of the project.

Whilst these steps will work to tackle specific barriers in the next stage of refinement of the project, practitioners also indicated several broader, systematic barriers to successfully implementing the intervention cannot be tackled as readily. For example, nearly all preschools had issues with staff shortages or staff sickness. Some preschools also reported a lack of cohesion between management and practitioners in the room or a free play-focused ethos which made it difficult to find time to carry out activities. It is also important to note that, while it was made clear that the intervention activities were primarily aimed at children who were due to start school the following year, including younger children in activities was often a priority for practitioners because of (positive) integration across age groups within their setting. When designing interventions at preschool, researchers should be aware of these broader barriers and work to ensure that the programme design is flexible to the differing needs of different preschools. Although setting-level differences in characteristics and implementation of the intervention may have had an impact upon outcomes, it was not possible to reflect on this due to the small sample size at the setting level (N = 8 intervention settings). Future evaluation with a greater number of settings is not planned for 2023 – 2025, funded by the Education Endowment Foundation and The Stronger Practice Hubs. It will allow for a more in-depth understanding of when and how the intervention is most effective.

From the point of view of efficacy, the next step needed is of course to test whether the current preliminary benefits of The ONE replicate on a much larger sample of children, particularly because the current sample of disadvantaged children was small. Previous findings suggest that individual differences in EF mediate the relationships between economic disadvantage and early numeracy (Blakey et al., 2020), so that disadvantaged children with greater opportunities to practice EF embedded in numeracy are likely to benefit most from this exposure. A future large scale trial is necessary to test this hypothesis with sufficient statistical power and is underway. Finally, if attitudes and emotions towards numbers already relate to numeracy outcomes, it is critical to continue fun, play based and engaging activities that foster an enjoyment of early numeracy, as attempted in The ONE.

## References

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences, 21,* 133–137. https://doi.org/10.1016/j.lindif.2010.09.013

Batchelor, S., Keeble, S., & Gilmore, G. (2015). Magnitude representations and counting skills in preschool children. *Mathematical Thinking and Learning, 17(2-3)*, 116-135.

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. PloS one, 9(11), e112393. https://doi.org/10.1371/journal.pone.0112393

Blakey, E., Matthews, D., Cragg, L., Buck, J., Cameron, D., Higgins, B., Pepper, L., Ridley, E., Sullivan, E. & Carroll, D.J. (2020). The Role of Executive Functions in Socioeconomic Attainment Gaps: Results from a randomised control trial. *Child development, 91(5),* 1594-1614, https://doi.org/10.1111/cdev.13358 .

Cahoon, A., Gilmore, C., Simms, V. (2021). Developmental pathways of early numerical skills during the preschool to school transition. *Learning and Instruction, 75.* https://doi.org/10.1016/j.learninstruc.2021.101484.

Coolen, I., Merkley, R., Ansari, D., Dove, E., Dowker, A., Mills, A., Murphy, V., von Spreckelsen, M., & Scerif G. (2021). Domain-general and domain-specific influences on emerging numerical cognition: Contrasting uni-and bidirectional prediction models. *Cognition, 215. https://doi.org/10.1016/j.cognition.2021.104816.*

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34 (02),* 819B.

Department for Education. (2020). *Development matters: Non-statutory curriculum guidance for the early years foundation stage.* London: Crown copyright.

Dietrich, J., Schmiedek, F., & Moeller, J. (2022). Academic motivation and emotions are experienced in learning situations, so let's study them. Introduction to the special issue. Learning and Instruction, 81. https://doi.org/10.1016/j.learninstruc.2022.101623

Dowker, A., Cheriton, O., Horton, R., & Mark, W. (2019). Relationships between attitudes and performance in young children's mathematics. Educational Studies in Mathematics, 100(3), 211-230. https://doi.org/10.1007/s10649-019-9880-5

Dowling, K., & Barry, M. M. (2020). Evaluating the Implementation Quality of a Social and Emotional Learning Program: A Mixed Methods Approach. I*nternational journal of environmental research and public health,* 17(9), 3249. https://doi.org/10.3390/ijerph17093249

Elliott, C.D., & Smith, P. (2011). *BAS3 British Ability Scales Technical Manual.* GL assessment.

Ganley, C. M., & McGraw, A. L. (2016). The Development and Validation of a Revised Version of the Math Anxiety Scale for Young Children. Front Psychol, 7, 1181. https://doi.org/10.3389/fpsyg.2016.01181

Hawes, Z., Merkley, R., Stager, C.L., & Ansari, D. (2021). Integrating numerical cognition research and mathematics education to strengthen the teaching and learning of early number. *British Journal of Educational Psychology, 91(4),* 1073-1109. https://doi.org/10.1111/bjep.12421

Hanner, E., Braham, E. J., Elliott, L., & Libertus, M. E. (2019). Promoting math talk in adult–child interactions through grocery store signs. Mind, Brain, and Education, 13(2), 110-118.

Howard, S. J., Vasseleu, E., Neilsen-Hewett, C., & Cliff, K. (2018). Evaluation of the Preschool Situational Self-Regulation Toolkit (PRSIST) Program for Supporting children's early self-regulation development: study protocol for a cluster randomized controlled trial. *Trials*, *19*(1), 64. https://doi.org/10.1186/s13063-018-2455-4

Howard, S. J., Vasseleu, E., Batterham, M., & Neilsen-Hewett, C. (2020). Everyday Practices and Activities to Improve Pre-school Self-Regulation: Cluster RCT Evaluation of the PRSIST Program. Frontiers in psychology, 11, 137. https://doi.org/10.3389/fpsyg.2020.00137

Howard, S.J , Neilsen-Hewett, C., de Rosnay, M., Melhuish, E.C., & Buckley-Walker, K, (2021). Validity, reliability and viability of pre-school educators' use of early years toolbox early numeracy. *Australasian Journal of Early Childhood, 47(2),* 92–106.

Howard, S.J., & Melhuish, E.C. (2017). An Early Years Toolbox (EYT) for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment, 35(3), 255-275.*

Hughes, C., Daly, I., Foley, S., White, N. and Devine, R. T. (2015). Measuring the foundations of school readiness: Introducing a new questionnaire for teachers – The Brief Early Skills and Support Index (BESSI). *British Journal of Educational Psychology, 85*, 332-356.

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE 8(7)*: e67918. https://doi.org/10.1371/journal.pone.0067918

Obersteiner, A. (2018). Multiple pathways between affect and mathematical competence in young children—commentary on the studies in the Special Issue. Educational Studies in Mathematics, 100(3), 317-323. https://doi.org/10.1007/s10649-018-9853-0

Scerif, G., Blakey, E., Gattas, S., Hawes, Z., Howard, S., Merkley, M., O'Connor, R., & Simms, V. (under review). Making the executive 'function' for the foundations of mathematics: The need for explicit theories of change for early interventions.

Siraj, I., Kingston, D., & Melhuish, E. (2015). *Assessing quality in early childhood education and care. Sustained shared thinking and emotional wellbeing (SSTEW) Scale for 2–5 year olds provision.* London: UCL and IOE Press.

Snijders, Tom A.B., and Bosker, Roel J. (2012). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, second edition. London etc.: Sage Publishers.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36(2),* 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1,* 297-301.

## Appendix

Note. The Figures below are examples of some of the questions in each subscale within the Child Attitudes Questionnaire – Preschool. Details of the questionnaire will be provided in the manuscript.


Figure A1. Example of the worry about numbers question within the Emotions subscale.


Figure A2. Number specific example question from the Attitudes subscale.


Figure A3. Example question from the General Motivation subscale.

Note. The Figures below plot estimated marginal means for numeracy and executive variables at Time 2, after controlling for Time 1 differences and modelling the effects of EYPP eligibility for children in the BAU arm and the Intervention (high adherence) arm.

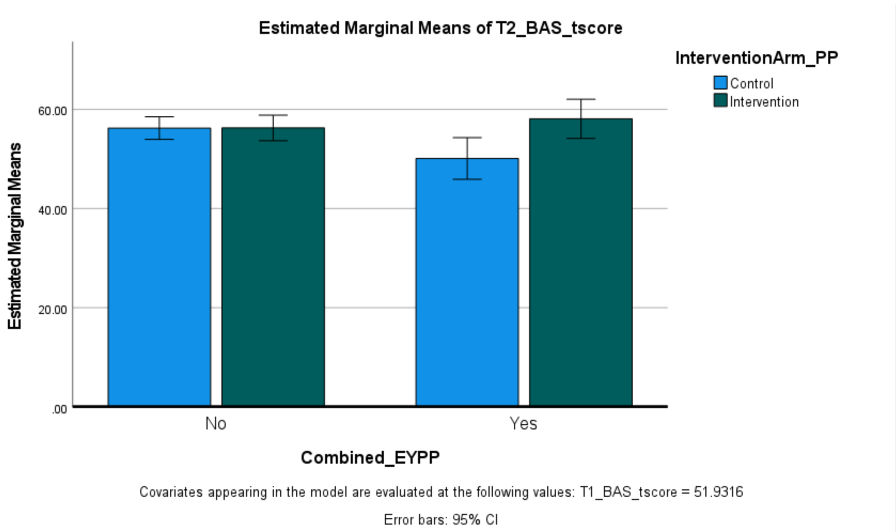



Figure A4. BAS tscores at Time 2, controlling for Time 1 scores.

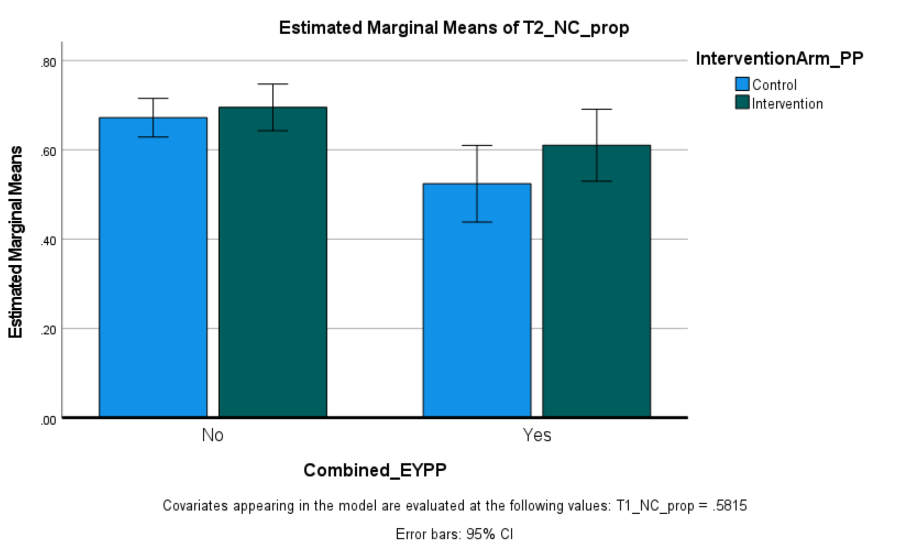



Figure A5. Digit Comparison scores (proportion correct) at Time 2, controlling for Time 1 scores.

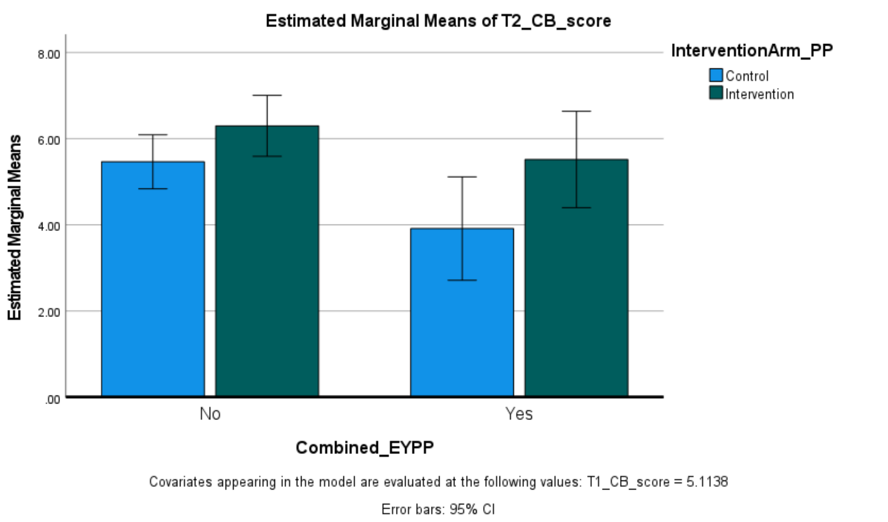



Figure A6. Corsi Block scores at Time 2, controlling for Time 1 scores.

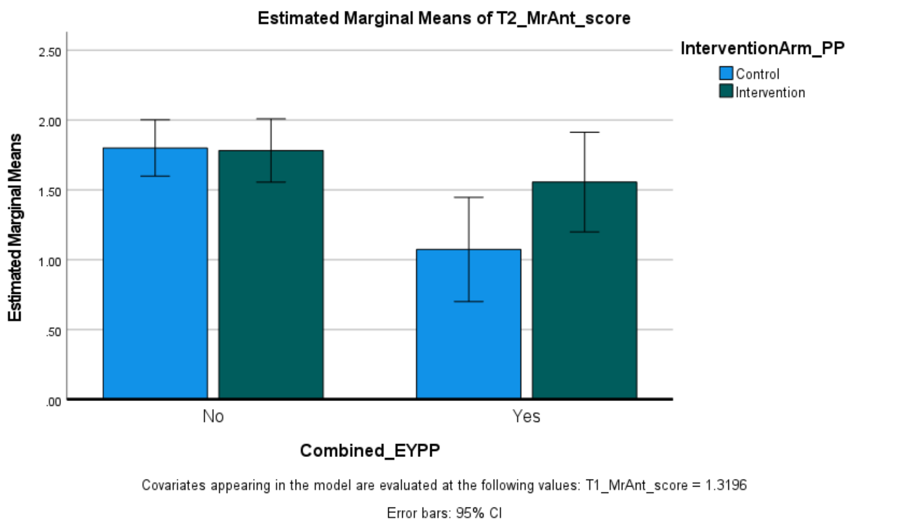



Figure A7. Mr Ant latent scores at Time 2, controlling for Time 1 scores.

Note. The Figures below plot mean scores on each measure at Time 1 and Time 2 in Control and both Intervention Groups (high and low adherence). On all plots, dots depict children's individual scores. Figures in blue represent numeracy variables, Figures in orange represent EF variables.

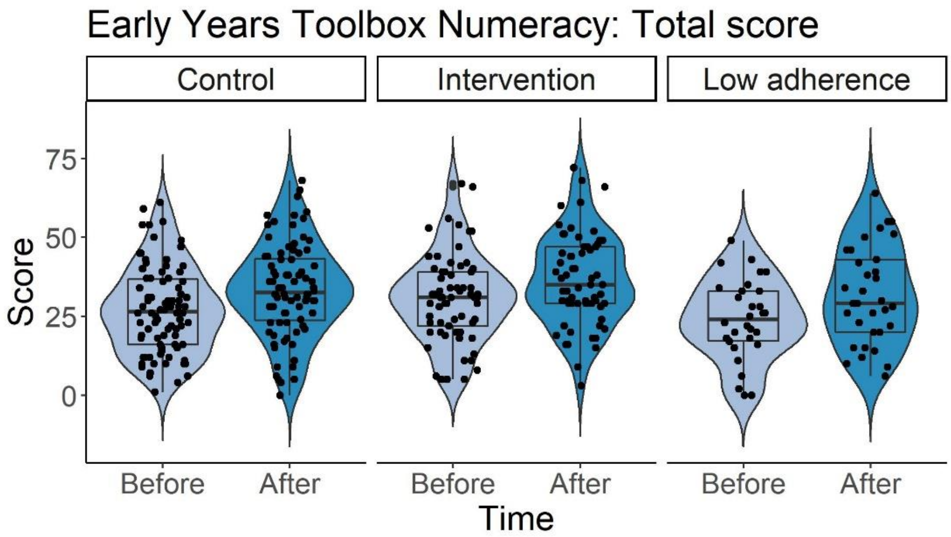



**Figure A8. EYTN total score.**

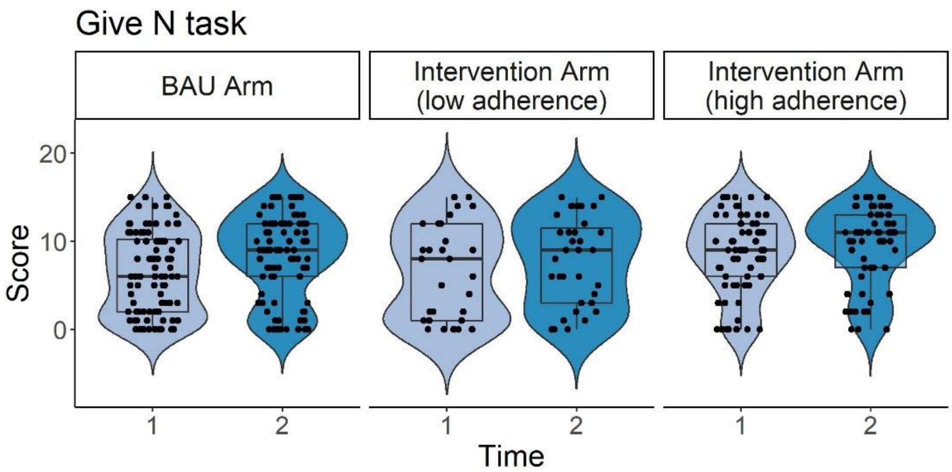



**Figure A9. Give N score.**

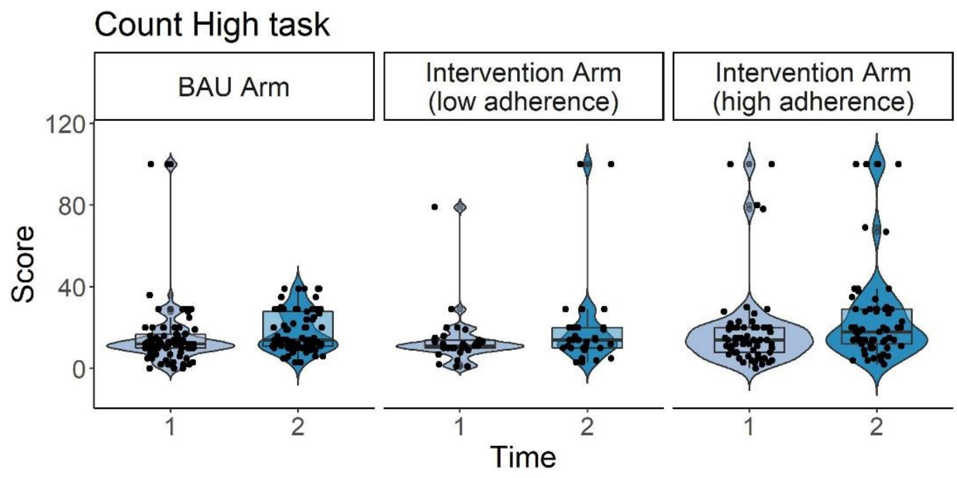



**Figure A10. Count high score.**

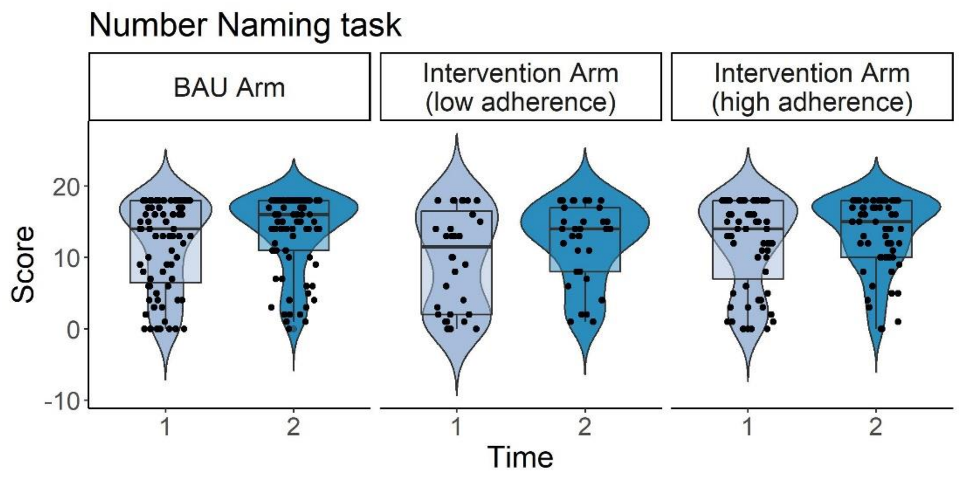



**Figure A11. Number Naming score.**

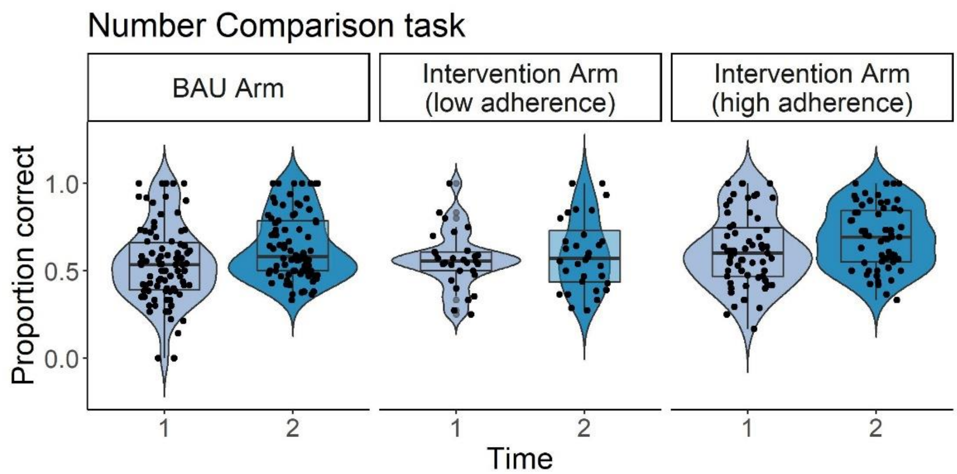



**Figure A12. Number comparison task: proportion of correct responses.**

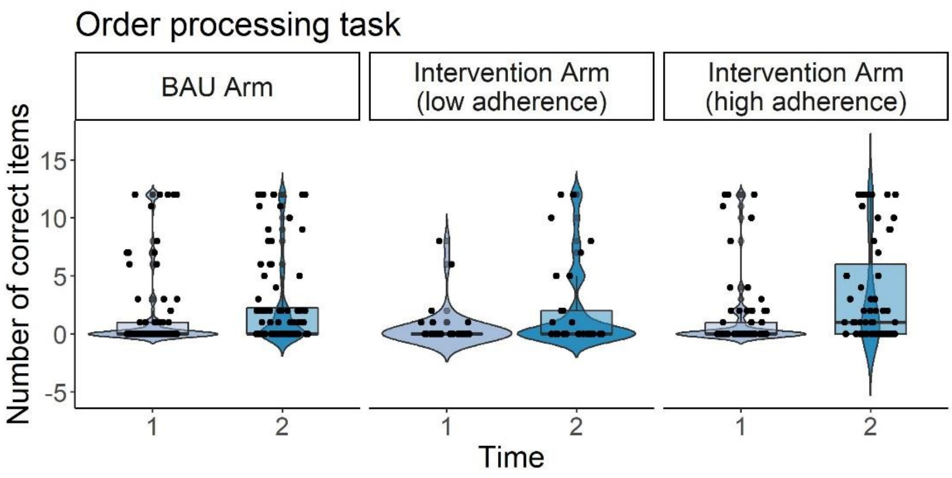



**Figure A13. Order processing task: score out of maximum 12.**

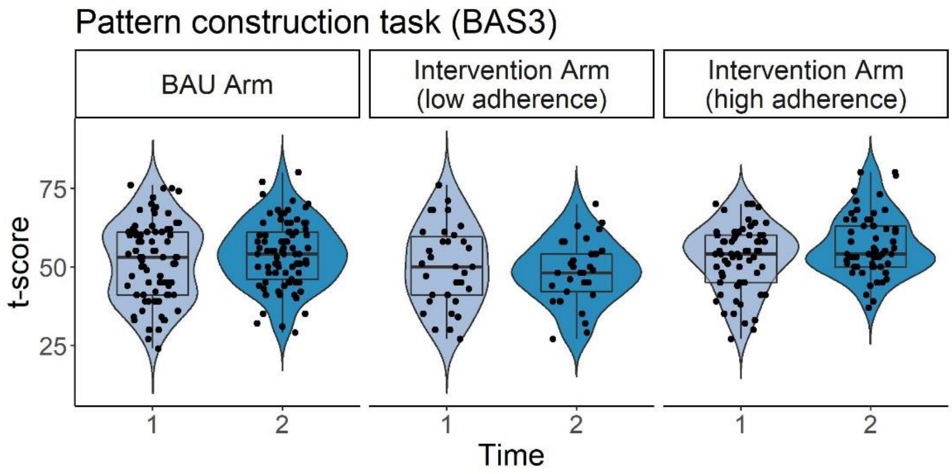



**Figure A14. British Ability Scale 3 pattern construction.**

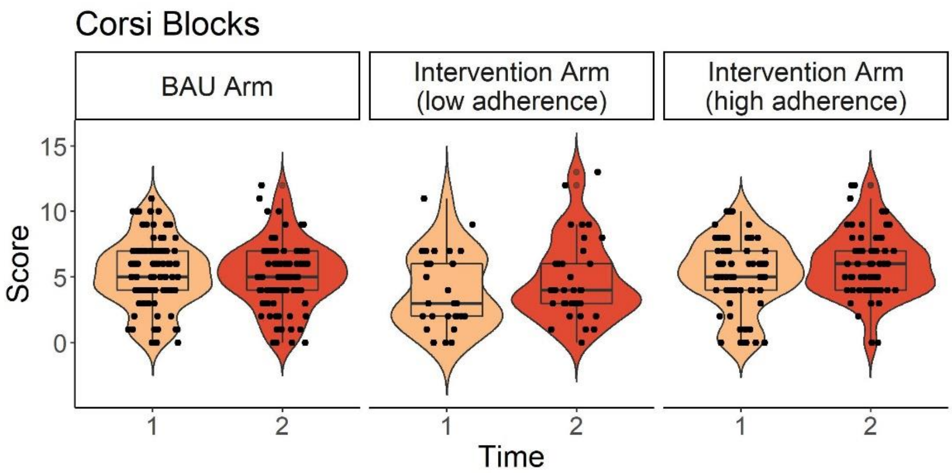



**Figure A15. Corsi blocks.**

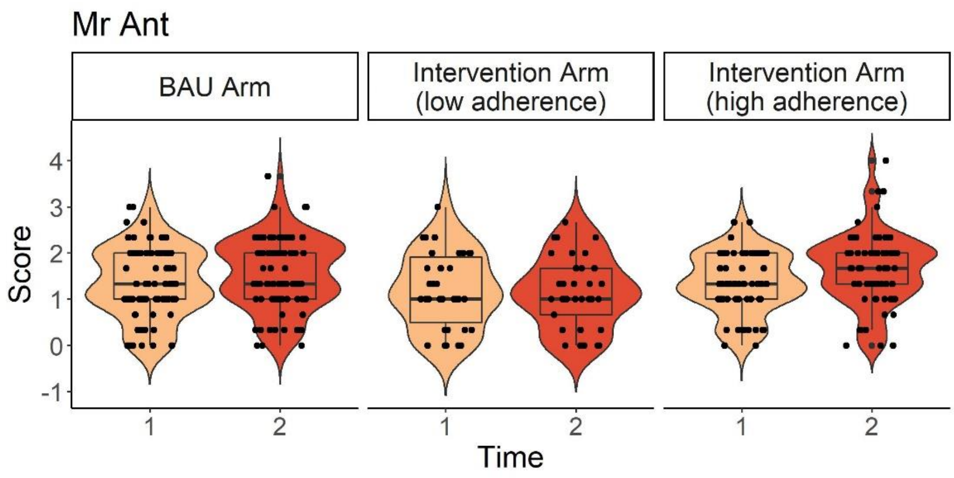



**Figure A16. Mr Ant.**

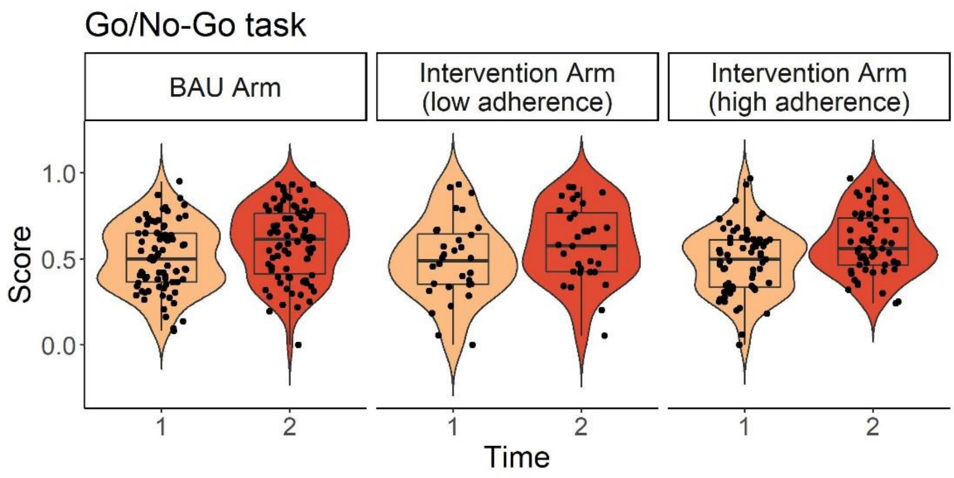



**Figure A17. Go – no go task (impulse).**

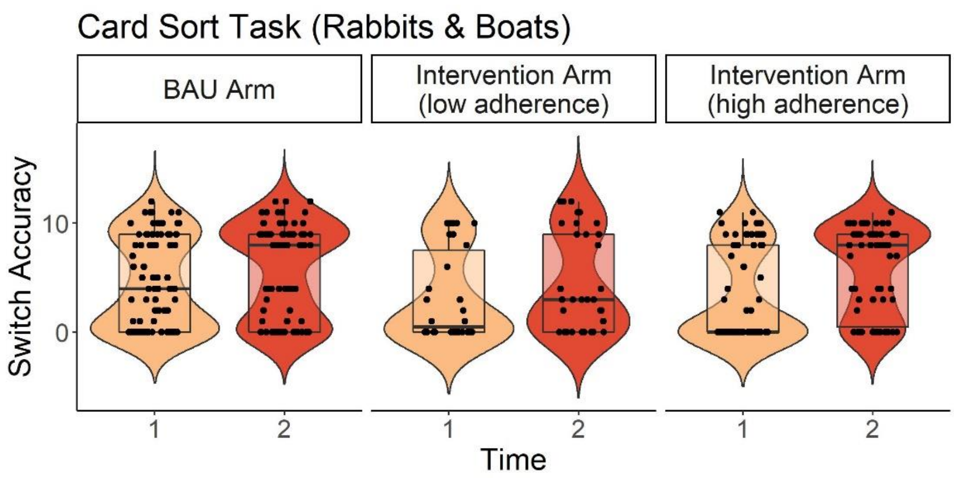



**Figure A18. Card sort task.**

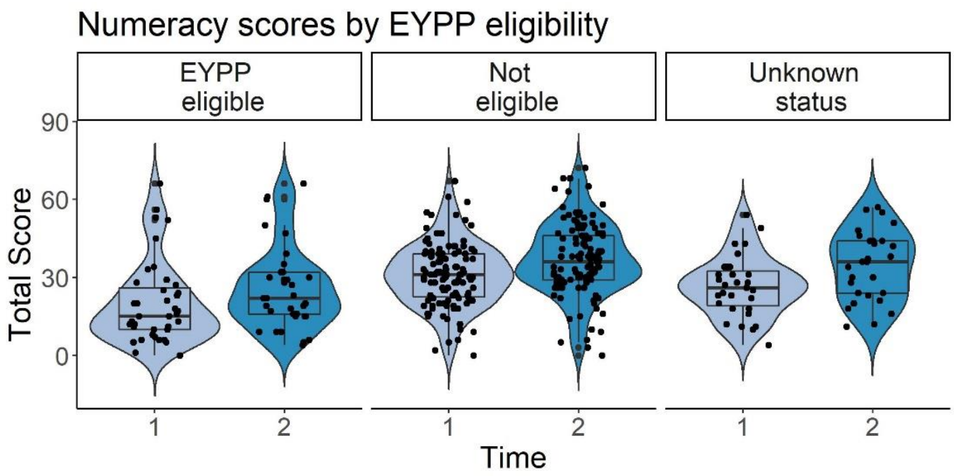



**Figure A19. EYTN scores of children with and without EYPP eligibility (NA category are children for whom this information was not available).**

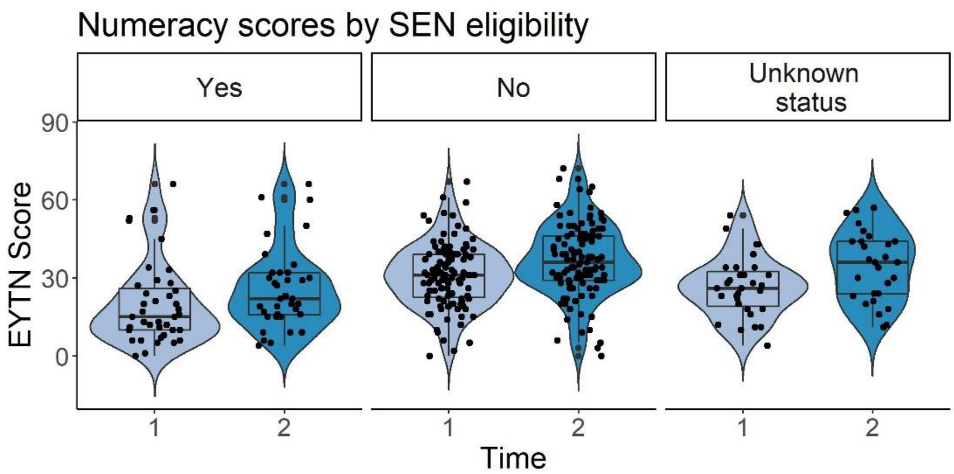



**Figure A20. EYTN scores of children with and without SEN (NA category are children for whom this information was not available).**

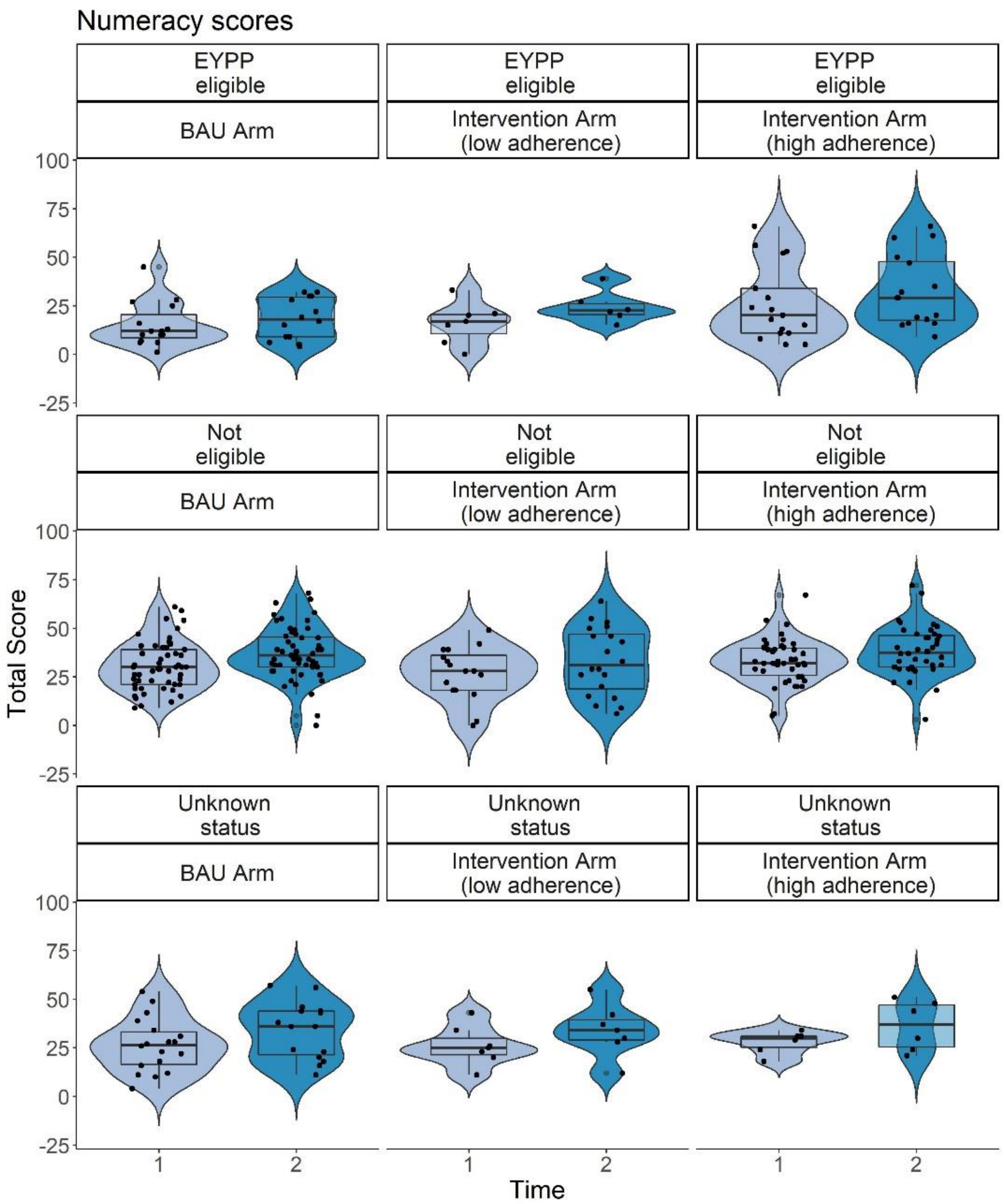
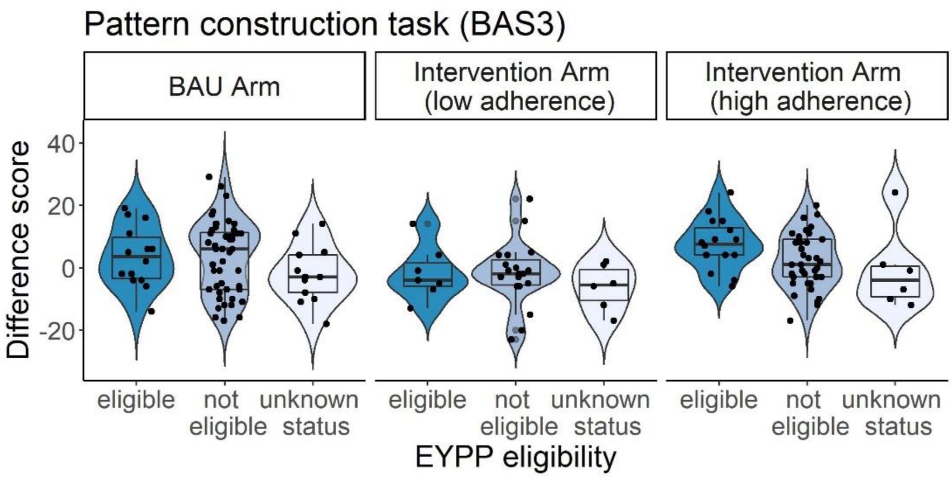
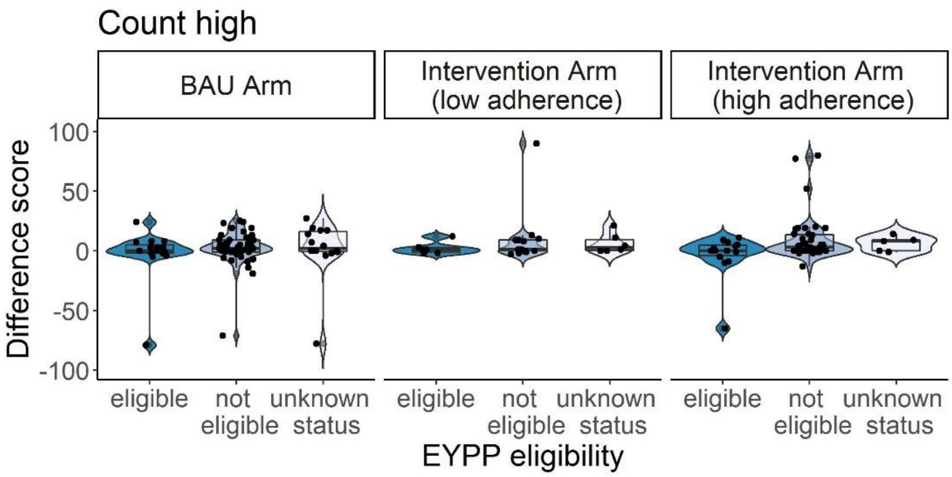
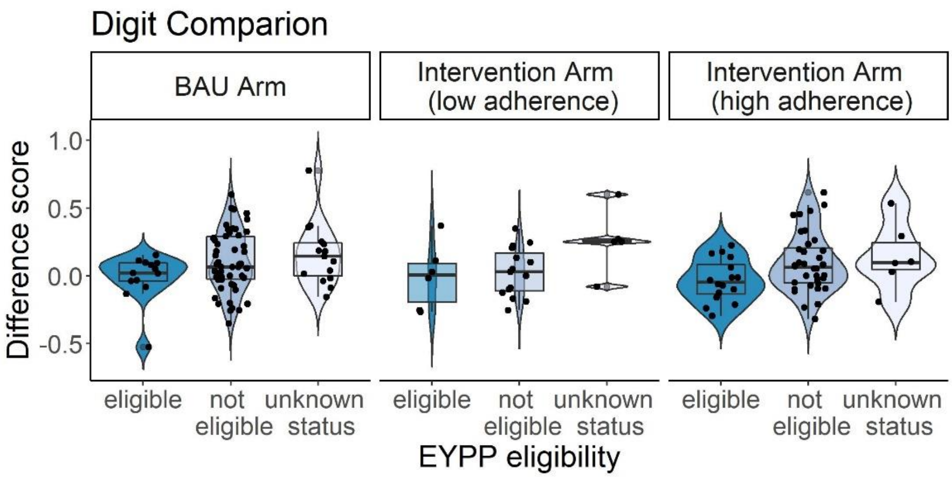


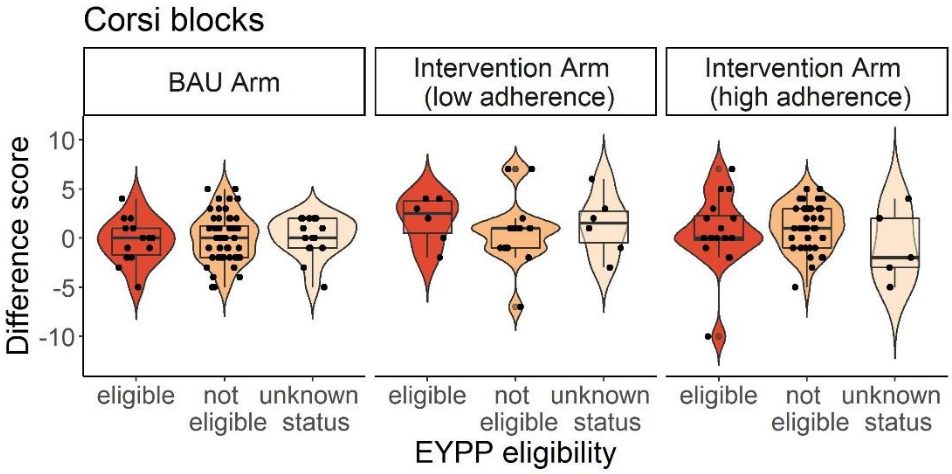

**Figure A21. Combined plot: EYTN scores in EYPP eligible and non-eligible children in Control, Intervention and Low Adherence groups.**

**Figure A22. Difference scores across measures for EYPP eligible and non-eligible children in Control, Intervention and Low Adherence groups.**




Corsi blocks
BAU Arm
Intervention Arm (low adherence)
Intervention Arm (high adherence)
Difference score
10
5
0
-5
-10
eligible
not eligible
unknown status
EYPP eligibility

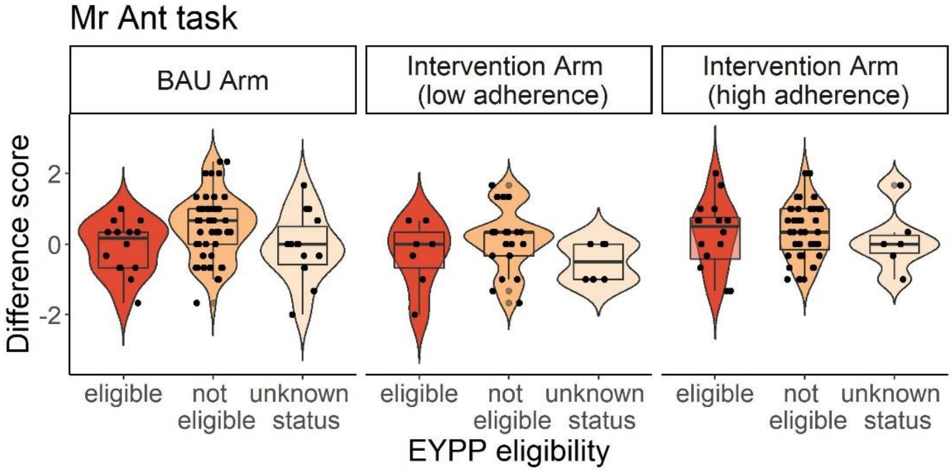

Mr Ant task
BAU Arm
Intervention Arm (low adherence)
Intervention Arm (high adherence)
Difference score
2
0
-2
eligible
not eligible
unknown status
EYPP eligibility

**Table A1: SSTEW maths subscale: counting/cardinality indicator**

**Counting/Cardinality**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 1.1 No use of numbers or counting is observed during the session<br><br>1.2 Counting is always led by the educator, with no opportunities for the children to join in | | 3.1 The children join in with counting, even if as a whole group activity<br><br>3.2 Educators incorporate counting into routines (e.g. register, snack time)<br><br>3.3 Activities, songs or books are chosen to include counting elements* | | 5.1 Most children are individually encouraged to join in with counting<br><br>5.2 Educators use pointing, fingers or props to aid children's understanding of counting<br><br>5.3 Counting is demonstrated in a variety of settings and using a variety of items (e.g. children, food items, pictures in a book) | | 7.1 Educators show an understanding of one-to-one correspondence by encouraging children to count slowly and point to one object at a time<br><br>7.2 Educators show an understanding of the cardinal principle, making sure to ask children "How many are there?" after counting or repeating the final number in a counting sequence.<br><br>7.3 Educators adapt counting activities to the abilities of different children, making sure that everyone is suitably challenged |

3.3 You may not see evidence of this. You could ask **question:** "When do you practice counting with the children?"

**Table A2: SSTEW maths subscale: shapes/spatial awareness indicator**

**Shapes and spatial awareness**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 1.1 Incorrect terminology is used to talk about shapes<br><br>1.2 Incorrect terminology is used to talk about the world around us | | 3.1 Educators talk about shapes using correct terminology (circle, triangle, square etc.)*<br><br>3.2 Activities and free play items are chosen to allow children to explore shapes (e.g. blocks, cut out shapes, playdoh ).<br><br>3.2 Educators talk about the world around us using correct terminology (e.g. under, on, over) | | 5.1 Shapes are mentioned outside of specific shape activities.<br><br>5.2 Shapes are described or compared to each other, either visually or verbally.*<br><br>5.3 Children are encouraged to use spatial language and pronouns, instead of simply pointing or saying "there". | | 7.1 Activities are chosen to encourage children to use spatial language (e.g. over, under, on) and educators actively encourage children to use this language.<br><br>7.2 Educators encourage children to think more deeply about shapes and spaces. For example, comparing shapes, estimating distances or planning a route)*<br><br>7.3 Shape and spatial language is carefully chosen at a level just above the children's understanding. |

3.1 , 5.2, 7.2 You may not witness any shape activities. **Question:** "How do you support the children to learn about shapes? Can you give me example activities?"

**Table A3: SSTEW maths subscale: order/patterning indicator**

**Order and Patterning**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 1.1 "Orderable" qualities of materials are never discussed with children (e.g. size, shape, weight).<br><br>1.2 Educators use language that is beyond children's comprehension to talk about patterning or order without offering any explanation | | 3.1 Educators use language related to order (e.g. first, second, third, after, before) and/or comparative adjectives (bigger, smaller, longer)<br><br>3.2 Children are encouraged to build or extend simple patterns*<br><br>3.3 Children are encouraged to put items in order or compare sizes of items | | 5.1 A rich variety of ordering language is used across different contexts (e.g. first/second/third, big/bigger/biggest).<br><br>5.2 Exploration of ordering and pattern are encouraged outside of specific order / pattern activities | | 7.1 Children supported to understand the concepts behind ordering language. This may be through directed questions, incorporation of ordering into daily routines or visual reminders of what ordering language means.<br><br>7.2 Educators make efforts to link patterning and order to children's day-to-day experiences or familiar stories. |

3.2 You may not witness any patterning games. **Question:** "Do you practice patterning? Can you show me what you?"

**Table A4: SSTEW maths subscale: number knowledge indicator**

**Number knowledge**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 1.1 Numbers are never used outside of counting in the classroom<br><br>1.2 Although there may be number displays in the classroom, they were made by adults and are never used in any child activities.* | | 3.1 Educators use visual numbers (or number representation like dots) in some games and activities.*<br><br>3.2 There are materials available that allow children to explore numbers during free play (e.g. relational rods, toy coins, hopscotch grid). | | 5.1 Numbers are represented in multiple ways around the preschool (e.g. digits, dots, fingers). Educators refer to these representations during activities.<br><br>5.2 Educators refer to numbers or draw children's attention to number representations in a natural way during non-number related games.<br><br>5.3 Children are given the opportunity to experiment with producing number symbols (e.g. digits, dots or tally marks).* | | 7.1 Educators support children's learning that numbers can be represented in many different ways. This may be through simultaneous use of different number representations, matching games or leading questions.<br><br>7.2 Educators show an understanding that children may need different types of number representations at different stages in the learning process, and use flexibly use number representations to support children's learning (e.g. using dots if a child doesn't know digits yet). |

1.2 If there are adult-made number displays and you do not observe them being used in class, you could ask **question:** "How was this display made? Do you ever use displays in activities?"

3.1; 5.3 You may not observe this, but you can mark this as a yes if you see evidence in classroom displays or children's work.

**Table A5. Activity frequency and ratings (out of 3), with selected comments taken from feedback sheets and interviews**

| Activity | No. times chosen | Mean rating (SD) | Selected comments from feedback sheet and interviews |
|---|---|---|---|
| What's the time Mr Wolf? | 22 | 3.00 (0) | *“Really fun, worked well with mixed age group”*<br>*“Big group outdoor games are always popular”* |
| See it, Build it, Check it | 18 | 2.50 (.797) | *“Awesome! Children were 'teachers' and could describe other 'mistakes'.”*<br>*“Tricky with colour and shapes. Only used 4-5 blocks each time to copy.”* |
| Special Hopscotch | 17 | 2.79 (.426) | *“The children did really well at taking turns and number negotiations.”*<br>*“...explored the idea of backwards”* |
| Secret Number | 17 | 2.92 (.277) | *“Children enjoyed this game”*<br>*“We used numbers 1 to 10. We put them in order then hid a number – asking the children what’s missing?”* |
| Can you draw this? | 16 | 2.73 (.467) | *“Good understanding of the shapes, but struggled in spatial awareness.”*<br>*“Challenged with more complex sequences”* |
| Number Robot | 16 | 2.86 (.363) | *“Great way to practice symbols, shapes, sizes”*<br>*“Fun to make the robot and a good game for exploring numbers”* |
| Special musical statues | 14 | 2.92 (.277) | *“Favourite! Worked well at big group time, with EF extension too.”*<br>*“We rolled a dice to decide on the number, which made the game more exciting”* |
| Counting Games | 13 | 2.55 (.522) | *“Some remembered rules and sequences, while others didn’t”*<br>*“Was hard to keep track of what was next.”* |
| Shape Sorter | 13 | 2.73 (.467) | *“Children compete to see who could fit the most shapes on paper.”*<br>*“The children enjoyed it but didn't know about shapes.”* |
| Secret Shadow Fingers | 12 | 2.40 (.843) | *“The children struggled to include their thumbs or to keep previous instructions in mind”*<br>*“Once explained it was okay. A little confusing.”* |
| Number Treasure Hunt | 12 | 2.82 (.404) | *“They enjoyed the challenge.”*<br>*“This was well liked, but we had a hard time getting the children to work together. Worked better to assign specific things for each child to look for.”* |
| Tower Challenge | 12 | 3.00 (0) | *“Enjoyed the challenge. Interesting when using objects of choice - didn't tell them what to use.”* |
| Numerical Rhythm | 11 | 3.00 (0) | *"Great way of practising pattern and order – this isn’t something we usually focus on. “* |

| | | | |
|---|---|---|---|
| Addition + subtraction ball | 9 | 2.50 (.756) | *“Too advanced for some of ours without support”* |
| Give me food | 8 | 2.71 (.489) | *“Kids enjoyed this lots. Started with 1-5 + dots and moved up to 6-10"* |
| Bigger Smaller | 8 | 2.50 (.755) | *“Added 'clap before you answer' - none could do this but could do an action after.”*<br>*“Good game to do over and over”* |
| War (card game) | 7 | 3.00 (0) | *“Tried lots of different versions.”*<br>*“Lots of anticipation, which held the children’s interest and made them feel relaxed. Adaptable to children’s number knowledge”* |
| What number am I? | 7 | 2.29 (.951) | *“Concept of asking Qs too tricky. Kept asking 'Is it 3?', 'Is it 5?'.”*<br>*“The children are getting better at this”* |
| Relational Rods: Roll + Build | 7 | 2.80 (.447) | *“Used numicons. Worked well, understood the dice and that numicons represented numbers.”* |
| Getting to know relational rods | 6 | 2.40 (.548) | *“More interested in names of colours than number property.”* |
| Make your own ruler | 6 | 2.75 (.500) | *“We extended the activity and it turned into a whole week of measuring things around the nursery school and making predictions about measurements.”*<br>*“The children picked this up better the second time”* |
| Little Biologists | 4 | 2.75 (.500) | *“A few children needed support to count corresponding objects.”*<br>*“Other outside activities were distracting”* |
| Make my train | 4 | 2.00 (1.00) | *"Children struggled and didn't like the rods. When shifted to patterns and challenge increased, children enjoyed more.”* |
| Mathematics Monsters | 4 | 2.75 (.500) | *“The children had fun with this. Lots of vocabulary for numbers/shapes/sizes.”*<br>*“It wasn’t very exciting or challenging for the older preschoolers”* |
| Free play | 2 | 3.00 (0) | *“Counting the animals in the book, using preposition: next to, under, behind.”* |

**Table A6.** Descriptive statistics for all secondary outcome measures, split by group (BAU control, a lower adherence group, and a higher adherence intervention group) and time (T1 and T2)

| Task | Time point | BAU Control | | | Intervention (high adherence) | | | Intervention (low adherence) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| **Numeracy Secondary Outcome Measures** | | | | | | | | | | |
| EYT Numeracy | T1 | 86 | 27.1 | 13.7 | 65 | 30.7 | 14.0 | 30 | 24 | 12.5 |
| | T2 | 84 | 33.4 | 14.8 | 62 | 36.6 | 14.5 | 33 | 31.7 | 15.1 |
| Give N total score (out of 18) | T1 | 88 | 6.47 | 4.66 | 65 | 8.60 | 4.51 | 29 | 6.66 | 5.26 |
| | T2 | 86 | 8.36 | 4.76 | 62 | 9.39 | 4.49 | 35 | 7.69 | 5.01 |
| Count high | T1 | 86 | 15.9 | 17.7 | 64 | 18.5 | 20.0 | 29 | 13.8 | 13.6 |
| | T2 | 81 | 17.9 | 9.78 | 61 | 25.4 | 23.9 | 34 | 19.4 | 21.4 |
| Number naming (out of 18) | T1 | 75 | 11.8 | 6.29 | 63 | 11.9 | 6.24 | 32 | 9.81 | 6.80 |
| | T2 | 82 | 13.6 | 5.37 | 62 | 13.4 | 5.12 | 33 | 12.2 | 5.40 |
| Digit comparison (proportion) | T1 | 87 | .541 | .221 | 62 | .615 | .208 | 29 | .562 | .160 |
| | T2 | 85 | .636 | .187 | 62 | .689 | .179 | 32 | .597 | .203 |
| Order processing (out of 12) | T1 | 81 | 1.75 | 3.64 | 64 | 1.42 | 3.23 | 32 | 9.81 | 6.80 |
| | T2 | 80 | 2.54 | 3.92 | 59 | 3.44 | 4.43 | 33 | 2.12 | 3.64 |
| BAS3 pattern construction t-score | T1 | 82 | 51.5 | 12.7 | 64 | 52.8 | 10.6 | 34 | 50.3 | 12.6 |
| | T2 | 81 | 54.0 | 10.6 | 61 | 57.0 | 9.82 | 33 | 48.3 | 9.96 |
| **Executive Functions Secondary Outcome Measures** | | | | | | | | | | |
| Corsi blocks total score | T1 | 87 | 5.18 | 2.65 | 65 | 5.22 | 2.69 | 29 | 4.00 | 2.84 |
| | T2 | 82 | 5.17 | 2.64 | 62 | 6.03 | 2.52 | 34 | 4.82 | 3.07 |
| EYT Mr Ant | T1 | 82 | 1.35 | .763 | 65 | 1.34 | .644 | 34 | 1.19 | .797 |
| | T2 | 85 | 1.55 | .768 | 62 | 1.65 | .799 | 33 | 1.58 | 1.05 |
| EYT Go/No-go impulse control | T1 | 79 | .509 | .195 | 65 | .484 | .193 | 32 | .495 | .226 |
| | T2 | 82 | .590 | .210 | 57 | .597 | .186 | 32 | .587 | .216 |
| EYT Card sort switch accuracy | T1 | 87 | 4.47 | 4.22 | 65 | 3.43 | 4.22 | 30 | 3.20 | 4.06 |
| | T2 | 85 | 5.65 | 4.30 | 62 | 5.76 | 4.04 | 35 | 4.37 | 4.48 |

**Table A7.** T2 scores for all tasks split by EYPP eligibility (yes/no/not available) and group (BAU control group / high adherence intervention / low adherence intervention)

| Measure | Eligible for EYPP? | BAU Control | | | Intervention (high adherence) | | | Intervention (low adherence) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| **Numeracy Secondary Outcome Measures** | | | | | | | | | | |
| EYT numeracy | No | 55 | 37.1 | 13.6 | 40 | 38.2 | 12.9 | 20 | 33.2 | 17.4 |
| | Yes | 14 | 18.4 | 10.7 | 16 | 32.6 | 18.6 | 6 | 24.3 | 8.19 |
| | N/A | 15 | 34.1 | 14.6 | 6 | 36.3 | 12.9 | 7 | 34.0 | 13.2 |
| Give N (out of 18) | No | 56 | 9.29 | 4.21 | 40 | 10.4 | 3.93 | 20 | 9.25 | 5.00 |
| | Yes | 15 | 4.13 | 5.30 | 16 | 7.00 | 5.40 | 8 | 4.28 | 3.58 |
| | N/A | 15 | 4.13 | 5.30 | 6 | 8.83 | 3.25 | 7 | 7.00 | 5.39 |
| Count high | No | 52 | 19.5 | 9.10 | 40 | 29.2 | 26.7 | 20 | 20.6 | 20.3 |
| | Yes | 15 | 11.3 | 7.02 | 15 | 17.2 | 17.1 | 7 | 10.9 | 5.70 |
| | N/A | 14 | 19.1 | 12.1 | 6 | 20.7 | 10.1 | 7 | 24.4 | 33.7 |
| Number naming (out of 18) | No | 54 | 14.3 | 4.92 | 40 | 14.0 | 4.61 | 19 | 14.2 | 4.14 |
| | Yes | 13 | 9.23 | 6.70 | 16 | 12.2 | 6.44 | 7 | 9.43 | 6.53 |
| | N/A | 15 | 14.9 | 3.94 | 6 | 13.0 | 4.69 | 7 | 9.57 | 6.08 |
| Digit comparison (proportion correct) | No | 56 | .668 | .197 | 40 | .707 | .170 | 19 | .623 | .212 |
| | Yes | 14 | .503 | .088 | 16 | .648 | .197 | 7 | .488 | .156 |
| | N/A | 15 | .640 | .168 | 6 | .683 | .196 | 6 | .642 | .229 |
| Order processing (out of 12) | No | 52 | 3.12 | 4.31 | 39 | 3.31 | 4.24 | 19 | 2.95 | 3.89 |
| | Yes | 13 | .769 | 1.42 | 14 | 3.71 | 5.20 | 7 | .286 | .756 |
| | N/A | 15 | 2.07 | 3.61 | 6 | 3.67 | 4.50 | 7 | 1.71 | 4.54 |
| BAS3 pattern construction t-score | No | 52 | 56.9 | 9.52 | 39 | 57.8 | 8.99 | 19 | 47.9 | 9.57 |
| | Yes | 15 | 44.5 | 8.03 | 16 | 57.0 | 12.2 | 7 | 45.0 | 9.92 |
| | N/A | 14 | 53.2 | 11.2 | 6 | 52.3 | 7.89 | 7 | 52.7 | 11.7 |
| **Executive Functions Secondary Outcome Measures** | | | | | | | | | | |
| Corsi blocks total score | No | 53 | 5.77 | 2.59 | 40 | 6.38 | 2.22 | 20 | 4.65 | 2.83 |
| | Yes | 14 | 2.71 | 2.33 | 16 | 5.31 | 3.24 | 7 | 3.43 | 1.72 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N/A | 15 | 5.33 | 1.68 | 6 | 5.67 | 2.16 | 7 | 6.71 | 4.31 |
| EYT Mr Ant | No | 55 | 1.74 | .702 | 40 | 1.78 | .722 | 19 | 1.12 | .803 |
| | Yes | 15 | .933 | .838 | 16 | 1.54 | .928 | 7 | 1.14 | .790 |
| | N/A | 15 | 1.44 | .613 | 6 | 1.06 | .772 | 7 | 1.24 | .738 |
| EYT Go/No-go impulse control | No | 52 | .596 | .218 | 39 | .591 | .187 | 18 | .591 | .250 |
| | Yes | 15 | .548 | .192 | 13 | .620 | .166 | 7 | .510 | .118 |
| | N/A | 15 | .612 | .205 | 5 | .581 | .256 | 7 | .654 | .213 |
| EYT Card sort switch accuracy | No | 56 | 6.39 | 4.11 | 40 | 6.40 | 3.66 | 20 | 3.75 | 4.87 |
| | Yes | 14 | 3.64 | 4.48 | 16 | 3.63 | 4.49 | 8 | 4.88 | 4.26 |
| | N/A | 15 | 4.73 | 4.35 | 6 | 7.17 | 3.71 | 7 | 5.57 | 4.20 |