Impact Evaluation of the Exploding Dots Program on the Development of Computational Thinking and Motivation for Learning Maths in Lower Secondary School Through an RCT

Education Endowment Foundation

Statistical Analysis Plan

Evaluator: Universidad de Oviedo
Principal Investigator: Luis J. Rodríguez Muñiz

PROJECT TITLE	Impact Evaluation of the Exploding Dots Program on the Development of Computational Thinking and Motivation for Learning Maths in Lower Secondary School Through an RCT
DEVELOPER	James Tanton's Global Math Project Implementation: MMACA - Museu de les Matemàtiques de Catalunya
EVALUATOR	Universidad de Oviedo
PRINCIPAL INVESTIGATOR	Luis J. Rodríguez Muñiz
PROTOCOL AUTHORS	Luis José Rodríguez Muñiz, Tania Iglesias
TRIAL DESIGN	Two-arm cluster randomised controlled trial with randomisation occurring at the school level
TRIAL TYPE	Efficacy
PUPIL AGE RANGE AND KEY STAGE	12-13 years old (1st-year Spanish secondary school students)
NUMBER OF SCHOOLS	86
NUMBER OF PUPILS	4750
PRIMARY OUTCOME MEASURE AND SOURCE	MDCT - Test Score (Mathematical Dimensions of Computational Thinking. Pending publication) [Santaengracia et al., 2024]
SECONDARY OUTCOME MEASURE AND SOURCE	IAM - Test Score - Attitudes Toward Mathematics Inventory [García et al., 2015]

SAP Version History

VERSION	DATE	REASON FOR REVISION
1.2 [*latest*]		
1.0 [*original*]	2 November 2023	

Table of contents

Introduction

The underlying hypothesis of this study is that arithmetic manipulation and number decomposition, as the bases of algebraic understanding, can positively influence the development of some dimensions of Computational Thinking (CT), particularly those not directly linked to algorithmics and programming.

The overall goal of this efficacy trial is to determine whether an intervention in mathematics teaching for 1st-year students of secondary school in Spain, which is centred around number sense and arithmetic, with a deep focus on place value, supported by Exploding Dots (ED), can foster the development of specific dimensions of computational thinking.

No prior studies closely resemble this research, as previous efforts have involved partial interventions that concentrated on teaching and evaluating the effectiveness of instructional approaches related to computational thinking or directly assessing students' computational thinking dimensions. These efforts [Román-González et al., 2017] mainly focused on the programming and algorithmic thinking aspects of computational thinking. The novelty of this study lies in exploring whether an indirect intervention aimed at strengthening number sense, which can potentially be extended to algebraic thinking, has a positive impact on skills that are considered fundamental for computational thinking.

Intervention

The intervention involved the implementation of an instructional path based on ED. ED is an innovative mathematics learning program designed for primary and secondary school students.

The intervention was conducted by the teaching staff in the 44 schools participating in the study, while an additional 42 schools served as the control group. The training consists of 10 in-person hours, delivered across 6 sessions over a weekend spanning 2 days. Additionally, there were 10 hours of asynchronous work, supplemented by ongoing support from the intervention team to assist with adapting to each school's specific needs.

Teachers were given access to the ED website, which provides various virtual manipulatives. A virtual platform was available to facilitate the course's follow-up. Materials associated with the different training sessions were made available and up to 2 online group sessions were scheduled.

Each teacher received a Teacher's Guide and a Student Workbook for implementing the first six ED units in the classroom. These materials were made available through a shared drive for teachers, who also had access to the ED website, where they could find various virtual manipulatives. The teaching phase occurred over one semester, from September 2023 to February 2024, with a frequency of 1 hour per week during regular mathematics class hours, totalling 17 weeks.

Design Overview

This was a two-arm cluster randomised controlled efficacy trial.

The schools were recruited in Catalonia, Aragon and Andalusia. Schools in Catalonia and Aragon were recruited through the Departments of Education of their respective regions. With a sample size of 80, we encountered challenges in finding a sufficient number of schools in the initial trial regions of Catalonia and Aragon, leading us to expand to Andalusia. This decision was influenced by the close collaboration with our partners from the HelloMath program and their significant connections with the Andalusian Government, which enabled us to achieve effective diffusion in the region.

Randomisation was conducted at the school level. Once the participant schools were determined, the selection of the two arms of the study was performed by simple randomisation with stratification based on the geographic area and type of school to ensure a balance between the intervention and control group. The variables considered for stratification included the geographical area in terms of the population size of the village/town/city in which the school was located (<50 000; 50 000-100 000; 100 000-250 000; >250 000 inhabitants) and the region (Catalonia, Aragon, Andalusia), the socioeconomic status of the families (low; medium-low; medium; medium-high; high), the type of school (state, foundation or private school) and the school's previous syllabus in computer science topics (yes; no).

Students in each school were treated anonymously; the project manager provided the schools with a method to identify the students using an alphanumeric code to ensure this anonymity. The codes remained consistent throughout the entire intervention, allowing for both individual and group measurements. The only student data collected in the study was age, gender, achievement level (1 – struggling, 2 – comfortable, 3 – excelling) and pretest and posttest scores. Students could only be identified by the school's administrators. The group allocation was also kept blind from the evaluation team.

The following research questions were considered:

1. Does the Exploding Dots learning program lead to improved maths and computer science skills in lower secondary students?

2. Does the Exploding Dots learning program enhance student motivation and enjoyment in learning mathematics?

The primary outcome is to measure the mathematical dimensions of students' computational thinking. This will be measured using an ad hoc instrument, designed by the evaluation team, the Mathematical Dimensions of Computational Thinking (MDCT) tool, implementing a shortened version in the pretest. The secondary outcome is to measure the attitudes towards mathematics, using an instrument extracted from the Attitudes Toward Mathematics Inventory *(*IAM), focusing on some of the dimensions measured in the survey.

Trial Design (including number of arms)		Two-arm randomised controlled trial at the school level
Unit of Randomisation		School
Stratification Variables (if applicable)		Geographic area (region and population size), socioeconomic status, type of school (private, foundation or state school), previous CS school syllabus(yes/no).
Primary Outcome	variable	Computational Thinking Dimensions
	measure (instrument, scale, source)	MDCT - Test Score (Mathematical Dimensions of Computational Thinking) Universidad de Oviedo
Secondary Outcome(s)	variable(s)	Attitudes towards mathematics, based on the following dimensions: lack of confidence in the future, perceived competence, perceived utility, intrinsic motivation, achievement motivation, lack of interest in mathematics, anxiety and feelings.
	measure(s) (instrument, scale, source)	Extract from IAM instrument (Spanish version) Universidad de Oviedo
Baseline for Primary Outcome	variable	Computational Thinking Dimensions
	measure (instrument, scale, source)	MDCT - Test score (Mathematical Dimensions of Computational Thinking) Universidad de Oviedo
Baseline for Secondary Outcome	variable	Attitudes towards mathematics, based on the following dimensions: lack of confidence in the future, perceived competence, perceived utility, intrinsic motivation, achievement motivation, lack of interest in mathematics, anxiety and feelings.
	measure (instrument, scale, source)	IAM instrument (Spanish version) Universidad de Oviedo scores

Sample Size Calculations Overview

The power analysis presents estimated Minimum Detectable Effect Sizes (MDES) for the primary outcome (MDCT Test). We considered a sample of 86 schools (44+42), providing a potential sample size of 5533 students (2722+2811). There were 4750 completed student responses (2424+2326). In comparison to other studies on Computational Thinking (CT), our study's sample size is notably large (e.g., the CT test study involved 24 schools).

		Protocol OVERALL	Randomisation OVERALL
Minimum Detectable Effect Size (MDES)		0.151	0.129
Pretest/Posttest Correlations	level 1 (pupil)	0.5	0.5
	level 2 (class)		
	level 3 (school)	0.5	0.5
Intracluster Correlations (ICCs)	level 2 (class)		
	level 3 (school)	0.05	0.05
Alpha		0.05	0.05
Power		0.8	0.8
One-Sided or Two-Sided?		2	2
Average Cluster Size		15	15
Number of Schools	intervention	40	44
	control	40	42
	total	80	86
Number of Pupils	intervention	1200	2424
	control	1200	2326
	total	2400	4750

Calculations of MDES and correlations were obtained for this sample size of 86 (44+42) using the PowerUpR R package. Due to the lack of scientific literature on our specific problem, we applied the maximum uncertainty principle, setting correlations at 0.5. Alpha values were set at 5 %, a common threshold in statistical studies.

From the protocol to randomisation, we changed the randomisation level from class level to school level. This change was necessary as classes cannot be identified: flexible class assignments are common, with students often sharing teachers. Moreover, the distribution of students in classes in Spain is generally random, usually in alphabetical order and never based on academic records. Therefore, randomisation was conducted at the school level.

Analysis

Pretest and posttest scores will be presented for each arm using mean and standard deviations, along with graphical representations. The results of the intervention between the experimental group and the control group will be compared, expressing the impact through the size of the standardised effect using Hedges' g with 95 % confidence intervals. A mixed-effects regression model will be constructed using the post-test score as the outcome variable, accounting for pupil-level and school-level effects. Gender, students' achievement levels (1 - struggling, 2 - comfortable, 3 - excelling) and students' previous experience in school activities related to computational thinking will be included in the model as fixed effects.

A pretest and posttest analysis of differences at both pupil and school levels, in terms of primary and secondary outcomes, will be conducted between the control and experimental groups. This analysis will be performed using an ANCOVA regression model with students' outcomes measured after the intervention.

Primary Outcome Analysis

The primary outcome is the measurement of decomposition, pattern recognition, generalisation and debugging dimensions of CT, using an ad hoc instrument designed for this study, called the Mathematical Dimensions of Computational Thinking (MDCT). The need to define an instrument stems from the aim of not focusing CT on programming skills, as previously available instruments mainly focused on programming and debugging. While programming is a core skill in CT, it is not the only one. Therefore, in line with the Bebras initiative, we aimed to focus on problem decomposition, modelling, data and other dimensions.

Questions in the pretest and posttests include the same contents to measure but have different formulations. The pre-test consisted of 17 single-choice items, each providing a stimulus and then offering four possible answers. The MDCT raw score is based on the number of correct answers. The instrument was administered through an online version (MS Forms). A mixed-effects model was constructed, using pretest and group, as well as the rest of the covariates as fixed effects. School was included as a random effect, a standard method for the analysis of cluster trials.

Model equation:

$Y_{ij} = \beta_0 + \beta_1 MDCT_{ij} + \beta_2 I_{Gj} + \beta_3 I_{Aj} + \beta_4 I_{Bi} + \beta_5 I_{Ci} + \beta_6 I_{Di} + \mu_i + \varepsilon_{ij}$

Y_{ij} posttest for j^{th} student in i^{th} cluster (school)

$MDCT_{ij}$ pretest for j^{th} student in i^{th} cluster (school)

I_{Gj} indicator variable for intervention/control group of j^{th} student

I_{Aj} indicator variable for gender of j^{th} student

I_{Bi} indicator variable for previous training in i^{th} cluster (school)

I_{Ci} indicator variable for population size in i^{th} cluster (school)

I_{Di} indicator variable for social economic status in i^{th} cluster (school)

μ_i random effect in i^{th} cluster (school)

ε_{ij} residual term for j^{th} student in i^{th} school

Secondary Outcome Analysis

The secondary outcome is the measurement of affective dimensions of mathematics education and uses a subset of the Attitudes Toward Mathematics Inventory (IAM) in the Spanish version. The selected subset comprises 32 items on a 1-4 Likert scale, covering the following dimensions: lack of confidence in the future, perceived competence, perceived utility, intrinsic motivation, achievement motivation, lack of interest in mathematics, anxiety and feelings. The IAM score consists of the addition of the individual scores of the items.

Model equation:

$Z_{ij} = \beta_0 + \beta_1 IAM_{ij} + \beta_2 I_{Gj} + \beta_3 I_{Aj} + \beta_4 I_{Bi} + \beta_5 I_{Ci} + \beta_6 I_{Di} + \mu_i + \varepsilon_{ij}$

Z_{ij} posttest for j^{th} student in i^{th} cluster (school)

IAM_{ij} pretest for j^{th} student in i^{th} cluster (school)

I_{Gj} indicator variable for intervention/control group of j^{th} student

I_{Aj} indicator variable for gender of j^{th} student

I_{Bi} indicator variable for previous training in i^{th} cluster (school)

I_{Ci} indicator variable for population size in i^{th} cluster (school)

I_{Di} indicator variable for social economic status in i^{th} cluster (school)

μ_i random effect in i^{th} cluster (school)

ε_{ij} residual term for j^{th} student in i^{th} school

Subgroup Analyses

No subgroup analysis will be done.

Additional Analyses

There will be no additional analyses.

Longitudinal Follow-Up Analyses

There will be no longitudinal follow-up analyses.

Imbalance at Baseline

Characteristics of the recruited schools will be presented overall. Pupil characteristics (gender, age and achievement level) will be summarised descriptively by randomised group, both as randomised and as analysed in the primary and secondary analyses. Continuous measures will be reported as a mean and standard deviation (SD), while categorical data will be reported as a count and percentage. The unadjusted difference between groups on the pretests will be reported as a Hedges' g effect size with a 95 % confidence interval.

Missing Data

Missing data are inevitable in a follow-up randomised controlled trial. A multilevel mixed-effect logistic regression model will be run to assess for statistically significant predictors of missing primary outcome data (where 1=missing; 0=complete) including all available pupil and school-level baseline data, with group as fixed effects and school as a random effect. Missing values will be identified in each variable and the pattern and type of missingness studied. A missing rate of 5 % or less would not typically bias the primary impact estimates, regardless of the pattern of missingness (Schafer, 1999) and so, a complete-case analysis will be employed. If missing data result in an exclusion of 5 % of data or more, a multiple imputation technique will be used and a sensitivity analysis will be performed.

Compliance

We define compliance at the school level. According to the Theory of Change, we have identified teachers' attendance at training and their adherence to their individual intervention plans as the key elements of the intervention, assigning each a weight of 35 % in the final compliance index. Furthermore, three other aspects have been evaluated during the intervention, allocating an equal weight of 10 % to each of them: level of teachers' involvement, the proportion of students actively involved and the extent of the use of the provided materials.

The considered factors, therefore, are:

- Average attendance percentage of the school's teachers at training (obtained from the attendance control lists).
- Average score of the work plan developed by the school's teachers (obtained from the score awarded by the intervention team, on a 0-100 scale).
- Average score of the school's teachers regarding their commitment to the intervention plan (obtained through a questionnaire administered during the intervention, on a 0-100 scale).
- Average percentage of the school's students who actively participated in the intervention, excluding students who sabotaged the activity (obtained through a questionnaire administered during the intervention).
- Average percentage of use of the materials developed by the school's teachers (obtained through a questionnaire administered during the intervention).

The total score will consist of an index (J) constructed at the school level, as follows:
$J = .35\cdot(a + b) + .1\cdot(c + d + e)$
J ranges between 0 and 100 and it will provide a measurement of the overall compliance. An individualised study of schools that do not achieve a score of 80 in indicator J will be conducted. The objective is to identify the circumstances that may have led to this outcome, potentially resulting in their exclusion from the sample if deemed necessary. Furthermore, a descriptive analysis will be performed to compare schools that surpass the 80-score threshold with those that do not. This aims to detect possible hidden associations between the variables involved that could impact the primary and secondary outcomes of the project.

Intra-Cluster Correlations (ICCs)

The intra-cluster correlation coefficient (ICC) associated with the school for the pretest and posttest outcomes will be provided with 95 % confidence intervals (CIs), determined using the ICC function of the R performance library. The ICC represents the proportion of variance in each outcome that can be explained by the variation between clusters (i.e., schools):

$$ICC = \frac{Random\ Intercept\ Variance}{Total\ Variance} = \frac{Random\ Intercept\ Variance}{Random\ Intercept\ Variance + Residual\ Variance}$$

The school-level intra-cluster correlation coefficient for the posttest primary outcomes will be extracted from each mixed model, with 95 % CIs. The ICC associated with the school for the pretest primary scores will also be presented with a 95 % CI.

Effect Size Calculation

Effect sizes for both outcomes will be calculated by dividing the adjusted mean difference between the intervention and control group by the pooled variance obtained from the unconditional model. A 95 % CI for the effect size will be calculated by dividing the 95 % confidence limits for the adjusted mean difference by this same variance. Thus, for the primary outcome:

$$ES = \frac{(\overline{Y}_I - \overline{Y}_C)}{\sqrt{s^*}}$$

Where:

$(\bar{Y}_I - \bar{Y}_C)$ denotes de difference in means between the trial arms obtained from the model and s^* denotes the pooled unconditional variance from the unconditional model. The same method will be used for the secondary outcome (W instead of Y in the formula above).

References

Fernández, T. G., Kroesbergen, E. H., Pérez, C. R., Castro, P. G., & García, J. A. G. P. (2015). Factors involved in making post-performance judgments in mathematics problem-solving. *Psicothema*, *27*(4), 374-380.

García, T., Rodríguez, C., Betts, L., Areces, D., & González-Castro, P. (2016). How affec-tive-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences*, *49*, 25-31.

Román-González et al. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*, *72*, 678-691.

Santaengracia et al. (2024). MDCT - Test Score (Mathematical Dimensions of Computational Thinking). *Pending publication*

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8(1), 3-15.