# Arukay: ESSA Tier III Report

## Evaluating the Arukay Learning System: A Correlational Study of CT, SEL, and Empathy in K–12

Summary report produced by the

International Centre For EdTech Impact

By Natalia Hernandez, Prof Nicola Pitchford and Jamie Elsner

July 2025

# 1. Executive Summary

This report presents findings from a quasi-experimental study evaluating the impact of the Arukay learning system on high school students' computational thinking (CT), social-emotional learning (SEL), and empathy. Conducted in two public schools in Bogotá, Colombia, the study involved 229 students: 136 in the intervention group who received CT instruction through the Arukay learning platform and 93 in the control group who received standard CT instruction as per the national curriculum in Colombia, over a 12-week implementation period.

Students in the intervention group engaged with Arukay's curriculum-integrated technology sessions, including coding activities in EduBlocks and MakeCode, mindfulness practices, and design thinking methodologies. Outcomes were measured using validated instruments tailored for adolescents at the end of the 12-week implementation period, using online assessments delivered in class.

Results indicate that students exposed to Arukay demonstrated significantly higher CT scores and stronger SEL and empathy competencies than their peers in the control group. The most notable gains were observed in sequencing, emotional competence, peer support, and perspective-taking. While further research is needed to examine long-term effects, these findings suggest that Arukay contributes meaningfully to students' cognitive and emotional development.

Aligned with ESSA Tier III ("Promising Evidence") and EduEvidence SILVER criteria, the study supports Arukay's scalability and relevance as a tool for promoting equitable learning in under-resourced educational settings. Recommendations for future research include deeper exploration of CT–SEL relationships, the use of pre-post designs to establish causal pathways, and continued alignment with national SEL mandates.

# 2. Introduction

### Background and Rationale

Computational thinking (CT) is an essential 21st century skill that provides learners with a structured approach to problem-solving including breaking down complex problems, recognizing patterns, designing algorithms, and abstracting. CT is defined as "the conceptual foundation required to solve problems effectively and efficiently (i.e., algorithmically, with or without the assistance of computers) with solutions that are reusable in different contexts" (Shute et al., 2017, p. 151). The benefits of CT on student learning are well-documented across multiple systematic and meta-analytic reviews (Ezeamuzie & Leung, 2022; Montuori et al., 2024; Cheng et al., 2023; Kalelioglu et al., 2016). In general, research on CT has focused primarily on measuring CT concepts such as programming or computing within STEM domains (Tang et al., 2020).

To prepare learners for solving real-world problems, which are inherently multi-disciplinary involving social and ethical domains, it is important to understand transferable skills that are essential in both STEM and non-STEM subjects. One such skill, socio-emotional learning (SEL), supports students in developing empathy, perspective-taking, collaboration, and resilience during problem-solving tasks. According to the CASEL framework, SEL development includes five core competencies: self-awareness, social awareness, relationship skills, self-management, and responsible decision-making (CASEL, 2025). During computational problem-solving, CT and SEL skills may be mutually beneficial to support learners in recognizing and responding to different viewpoints and societal needs required for solution design. To support the broader goal of developing innovative and compassionate problem-solvers, it is important to understand the nature of the relationship between CT and SEL in educational contexts.

### Association between CT and SEL

Computational thinking often requires collaboration in teams, particularly when learners need assistance overcoming coding challenges such as debugging. Compared to CT only approaches, integrating CT with a SEL approach may create more holistic problem-solving skills that encourage creativity and efficiency while also enhancing social and emotional skills. Yet, despite the possible advantages of integrating CT and SEL in education programs, evidence of the relationship between CT and SEL is still emerging.

A recent study by Candeias et al. (2025) proposed a program for CT and SEL integration as part of the COMPUSEL educational project which aimed to enhance students SEL in the five key areas of the CASEL framework (i.e., self-awareness, social awareness, relationship skills, self-management, and responsible decision-making). The curriculum featured digital stories centered around socio-emotional skills and corresponding activities involving computational principles. The project has been implemented across five countries including Portugal, Greece, Poland, Romania, and Turkey (Sakellariou, 2023).

Candeias et al. (2025) measured the impact of the COMPUSEL curriculum and found significant improvements in students' SEL competencies in all five aspects, suggesting positive benefits of an integrated CT and SEL program to support students' problem-solving of social and interpersonal challenges. Qualitative results from teacher interviews reinforced the finding that students used CT principles when resolving interpersonal conflict. This aligns with a previous study which found that expression of empathy, such as concern for others or imagining how others may feel, informed the setup of CT algorithms (Sohr et al., 2023).

However, results also showed that the effects of the COMPUSEL project varied across countries, indicating the need to pay attention to the cultural context in which the intervention is situated. For example, educational policies, teacher professional development training, and school context may influence the implementation of SEL interventions across countries. The researchers suggested that further studies are needed spanning diverse educational settings to better understand how integrated CT and SEL interventions might be effective cross-culturally.

### Theoretical Rationale

The current study is situated in Columbia with Spanish-speaking students thus it addresses the research gap identified above by expanding understanding of the relationship between CT and SEL across different cultural contexts. Additionally, it builds upon a previously developed logic model, theory of change, and theory of action, in a report that was certified with a Bronze Badge in Efficacy by EduEvidence in May 2025, as meeting ESSA Tier IV standards. That report presented the rationale for expecting an association between CT and SEL and empathy, based on theories of cognitive, affective, and social-emotional development. The present study is designed to examine empirically this rationale through a correlational design, that explores the relationship between CT and SEL and empathy between a group of students receiving the Arukay CT learning system and a group of students receiving standard instruction in CT as specified by the national curriculum in Colombia.

### Objectives of the Study

This study aims to examine if participation in the Arukay learning system is associated with enhanced computational thinking (CT), socio-emotional learning (SEL), and empathy-related skills compared to standard educational practice in Colombia. It also seeks to explore if the Arukay learning system strengthens associations between these domains.

### Research Questions

The following research questions were addressed. Compared to standard practice:

1. Does Arukay improve CT skills?
2. Does Arukay improve SEL skills?
3. Does Arukay improve empathy-related skills?
4. Does Arukay enhance the association between CT and SEL?
5. Does Arukay enhance the association between CT and empathy-related skills?

# 3. Intervention Description

The Arukay intervention began on February 11, 2025, and concluded on May 23, 2025. It was implemented in four technology class streams: two from 10th grade and two from 11th grade, reaching a total of 145 students. Each stream was initially scheduled for 14 sessions; however, due to operational adjustments, students received between 10 and 12 sessions. The 10th-grade students participated in the *Arukay Coding 5* course, while the 11th-grade students followed the *MakeCode Arcade Video Game Design* course. Both programs aimed to foster computational thinking through hands-on programming activities, aligned with Arukay's pedagogical framework.

The intervention took place in a dedicated computer lab where each student had individual access to a computer and worked at their own pace. The sessions were led by their technology teacher, who followed Arukay's structured lesson guides. Each class followed a consistent structure:

- A short mindfulness session to begin the class,
- Presentation of the learning objective, and
- Hands-on practice through the Arukay platform, using personalized student accounts.

A member of the Arukay team was present during all sessions to observe and provide technical or instructional support as needed.

***Teacher Training***

Prior to implementation, the teacher participated in a training program led by Arukay, which covered:

- The pedagogical and didactic foundations of the Arukay methodology,
- Technical use of the learning platform, and
- In-depth training on the programming languages used in each course.

***Course Description: Arukay Coding 5***

Arukay Coding 5 is designed to strengthen students' computational thinking and problem-solving abilities using block-based Python programming in EduBlocks. Each session follows the Design Thinking framework, guiding students to empathize with a problem, define needs, ideate solutions, prototype their code, and test their projects iteratively. The course emphasizes the development of digital solutions related to sustainability, digital well-being, and emotional regulation. By the end of the course, students are expected to:

- Apply principles of computational thinking,
- Design and program interactive projects in EduBlocks,
- Work collaboratively and think critically,
- Demonstrate empathy in problem-solving,
- Communicate project ideas effectively, and
- Use Design Thinking to create innovative digital solutions.

The average course progress across students was 96%, reflecting a high level of engagement and completion of the course objectives.

***Course Description: MakeCode Video Game Design***

The MakeCode course introduces students to 2D game development through Microsoft's MakeCode Arcade platform. Each session allows students to experiment with core components of video game creation, such as sprites, animation, interactivity, and sound integration. As the course progresses, students incrementally apply programming and design concepts to develop playable games. Key learning objectives include:

- Designing and developing 2D arcade games using coding and computational thinking,
- Understanding game development concepts such as control logic, animation, game flow (start–play–end), and object interaction, and
- Enhancing creativity, logical thinking, and technical skills in a playful, exploratory environment.

# 4. Study Design

This study employed a correlational design with quasi-experimental elements. It included an intervention group who received instruction in CT through the Arukay learning system and a control group, who received standard CT class-based instruction. Both groups were drawn from public schools in Bogotá, Colombia.

The intervention group consisted of students from *Colegio La Bici*, where the Arukay learning system was implemented in four class streams of the technology course for 10th and 11th graders. The control group was drawn from *Institución Educativa Distrital El Recuerdo*, which also included students from the 10th and 11th grades.

Schools were selected based on specific criteria as requested by the funders of Arukay. These included being public institutions and having an adequately equipped technology classroom with the minimum technological requirements to support the Arukay learning system. The funders specifically requested that the intervention focus on public school settings to align with their social mission of promoting educational equity. By targeting schools serving historically under-resourced communities, the intervention aims to support scalable solutions that are not only innovative but also capable of "changing realities" and expanding access to quality education for all students.

**Ethical considerations**

This study followed ethical principles in accordance with national and institutional guidelines for research involving human participants in Colombia. Informed consent was obtained from parents or caregivers of students prior to participation. Students were informed that participation in the study was voluntary and that they could withdraw at any time without consequences. All data was anonymized to protect participant identities and was used exclusively for research purposes. In addition, school leaders granted permission for the study to take place at their institution after being fully informed of its purpose and methodology.

# 5. Measures and Instruments

Three domains were assessed in this study: computational thinking, socio-emotional learning, and empathy.

Computational thinking was measured using the *Computational Thinking Test* (CTt), developed by Universidad Rey Juan Carlos. This instrument evaluates core competencies such as abstraction, algorithmic thinking, and problem-solving, and has been widely used in school-based interventions focused on digital skills (Román-González, 2015).

Socio-emotional learning was assessed with the *Social-Emotional Health Survey-Secondary* (SEHS-S), a validated tool designed to capture strengths-based indicators such as self-regulation, school support, and optimism among adolescents (Riaño-Hernández et al (2025).

Empathy was evaluated using two complementary instruments: the *Adapted Empathy in Design Scale* (EMPA-D), which focuses on empathy in collaborative and design-based contexts (Drouet et al, 2024), and the *Basic Empathy Scale* (BES), which measures both affective and cognitive empathy in adolescents (Salas-Wright et al, 2013). The EMPA-D was adapted for use with high school students, as the original version was developed for adults working in professional design environments. The item wording was revised to better align with the context and language of secondary school students. The adapted version was reviewed by two pedagogical experts from Arukay to ensure age-appropriate phrasing and conceptual clarity for the target population.

# 6. Data Collection Procedures

Data collection was carried out during the last week of May 2025, specifically from May 26 to May 30. Prior to this, researchers met with teachers from both schools to explain the data collection protocol. The data collection process was conducted online during the schools' scheduled technology sessions. A staff member from Arukay was present on-site to support teachers, while researchers provided remote assistance as needed throughout the administration process.

To monitor participation in real time, researchers remained in contact with the teachers and the staff member of Arukay and used an anonymized list system to track response progress. Although the data collection was anonymous, each student was assigned a list number that allowed the team to identify missing responses without revealing student identities. This system made it possible to follow up promptly with teachers to encourage full participation and address any technical or logistical issues during the sessions.

A total of 229 students completed the online survey during the week allocated for data collection. Missing data was treated as attrition. Nine students enrolled in the intervention group were absent during data collection, so they were excluded from the analysis. Information regarding the number of students in the control group was not available, so attrition in that group could not be assessed. Additionally, some inconsistencies in the socioeconomic stratum variable were observed from the student responses which were addressed through data cleaning. Specifically, responses indicating they belonged to strata 4 through 6 were excluded after verification by school officials confirmed that all students belonged to strata 1 to 3.

# 7. Participants

The final sample consisted of 229 students, 136 from the intervention group, and 93 from the control group (see Table 1). All participants were enrolled in enrolled in grades 10 and 11, except for one participant in the intervention group who was in grade 9. The participants' ages ranged from 14 to 20 years, with most students aged 16 years (63 (46%) in the intervention group and 43 (46%) in the control group) or 15 years (35 (26%) and 13 (14%), respectively). The sample included a relatively balanced distribution across sex, with females representing 63 (46%) participants in the intervention group and 48 (52%) in the control group, and males representing 71 (52%) and 44 (47%), respectively. Additionally, one participant in each group identified as non-binary, and one participant in the intervention group preferred not to disclose their sex. In terms of socioeconomic status, most of the intervention group belonged to stratum 2 (n = 101, 74%), whereas the control group was primarily composed of students from stratum 1 (n = 67, 72%), reflecting differing economic profiles between the two groups. However, it is important to note that in the national classification system, strata 1 and 2 are considered part of the low-income range. Detailed sociodemographic characteristics of both groups are presented in Table 1.

*Table 1. Sociodemographic Characteristics of Participants by Study Group.*

| Variable | Category | Intervention Group | Control Group |
|---|---|---|---|
| Grade | 9th | 1 | 0 |
| | 10th | 74 | 50 |
| | 11th | 61 | 43 |
| Age | 14 | 8 | 1 |
| | 15 | 35 | 13 |
| | 16 | 63 | 43 |
| | 17 | 24 | 22 |
| | 18 | 6 | 11 |
| | 19 | 0 | 2 |
| | 20 | 0 | 1 |
| Sex | Female | 63 | 48 |
| | Male | 71 | 44 |
| | Non-binary | 1 | 1 |
| | Prefer not to say | 1 | 0 |
| Socioeconomic Stratum | 1 | 3 | 67 |
| | 2 | 101 | 20 |
| | 3 | 29 | 3 |

# 8. Data Analysis

To address the research questions, quantitative methods were applied. First, descriptive statistics were calculated for all key variables related to computational thinking, socio-emotional learning, and empathy.

Normality of the total scores and sub-scores was assessed using the Shapiro-Wilk test, which revealed that most variables were not normally distributed. As a result, non-parametric approaches were used for subsequent analyses. For group comparisons, the Wilcoxon rank-sum test (equivalent to the Mann-Whitney U test) was used instead of the parametric t-test. To examine associations between variables, Spearman's rank correlation coefficient was used in place of Pearson's Product Moment correlation.

# 9. Findings and Discussion

Table 2 presents descriptive statistics for all key measures of this study. Descriptions of the score calculations are presented in each subsection below.

*Table 2: Descriptive statistics for all measures*

| Score | Group | Min | Q1 | Median | Mean | Q3 | Max |
|---|---|---|---|---|---|---|---|
| Computanional Thinking | Intervention | 8.00 | 16.00 | 19.50 | 19.00 | 22.20 | 26.00 |
| | Control | 5.00 | 10.00 | 13.00 | 13.10 | 16.00 | 23.00 |
| Socio-Emotional Thinking | Intervention | 1.89 | 2.97 | 3.21 | 3.15 | 3.44 | 4.00 |
| | Control | 1.89 | 2.69 | 3.03 | 3.00 | 3.31 | 3.94 |
| Empathy (EMPA) | Intervention | 1.73 | 3.55 | 3.91 | 3.86 | 4.36 | 5.00 |
| | Control | 1.00 | 3.09 | 3.55 | 3.54 | 4.00 | 5.00 |
| Empathy (BES) | Intervention | 30.00 | 60.00 | 69.00 | 69.20 | 79.20 | 100.00 |
| | Control | 25.00 | 55.00 | 66.00 | 63.60 | 75.00 | 94.00 |

For all graphs reported below significance levels are given as * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

## Computational Thinking Scores

### Score calculation

Scores for the Computational Thinking Test were calculated by summing the number of correct responses across all test items to create a total score for each participant. In addition, three sub-scores were computed, corresponding to specific dimensions of computational thinking: sequencing, debugging, and completion. Each sub-score was based on the sum of correct responses to the items corresponding to that dimension.

**Results**

Shapiro-Wilk test revealed that most variables significantly deviated from a normal distribution ($p < .05$), particularly in the intervention group (see Appendix A). Only the Total CT Score in the control group met the assumption of normality. Due to these violations, non-parametric statistical tests were deemed most appropriate for comparative analyses between groups.

A significant difference was observed in the Total CT Score between the intervention school and the control school ($W = 10{,}534$, $p < 0.001$), confirming the positive impact of the Arukay learning system across the broader construct of computational thinking, as shown in Figure 1. Consistent with the overall results, statistically significant differences were found between the intervention school and the control school across all three subdimensions of CT, as shown in Figure 2. Specifically, the intervention group obtained higher scores than the control group in Completion ($W = 10{,}126$, $p < 0.001$), Debugging ($W = 9{,}600$, $p < 0.001$), and Sequencing ($W = 10{,}062$, $p < 0.001$). These findings indicate that students exposed to the Arukay learning system demonstrated better performance in fundamental CT skills than students receiving standard instruction in CT.

These results provide evidence in support of the first research question regarding computational thinking. Compared to standard practice, instruction through the Arukay learning system led to significantly higher median scores in all three CT components as well as in the overall CT measure. These findings suggest that the Arukay learning system is associated with measurable improvements in students' computational thinking abilities compared to standard instructional practice.

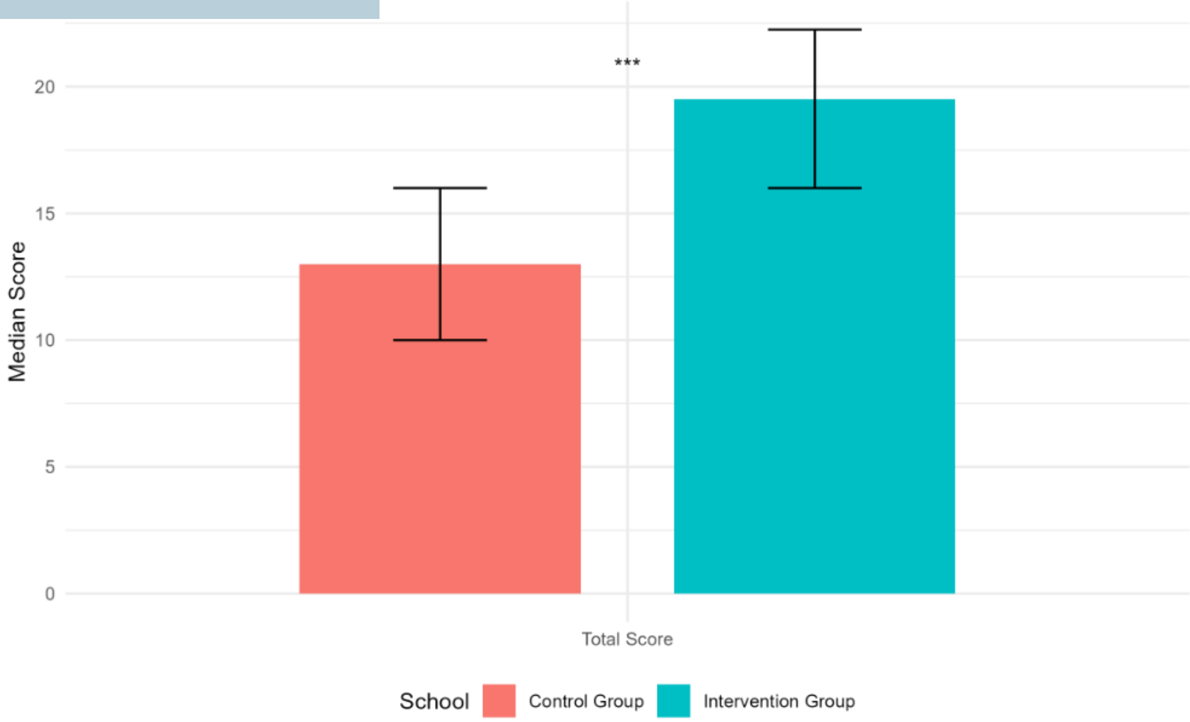


*Figure 1: Median total CT scores by school, including significance level of group difference.*

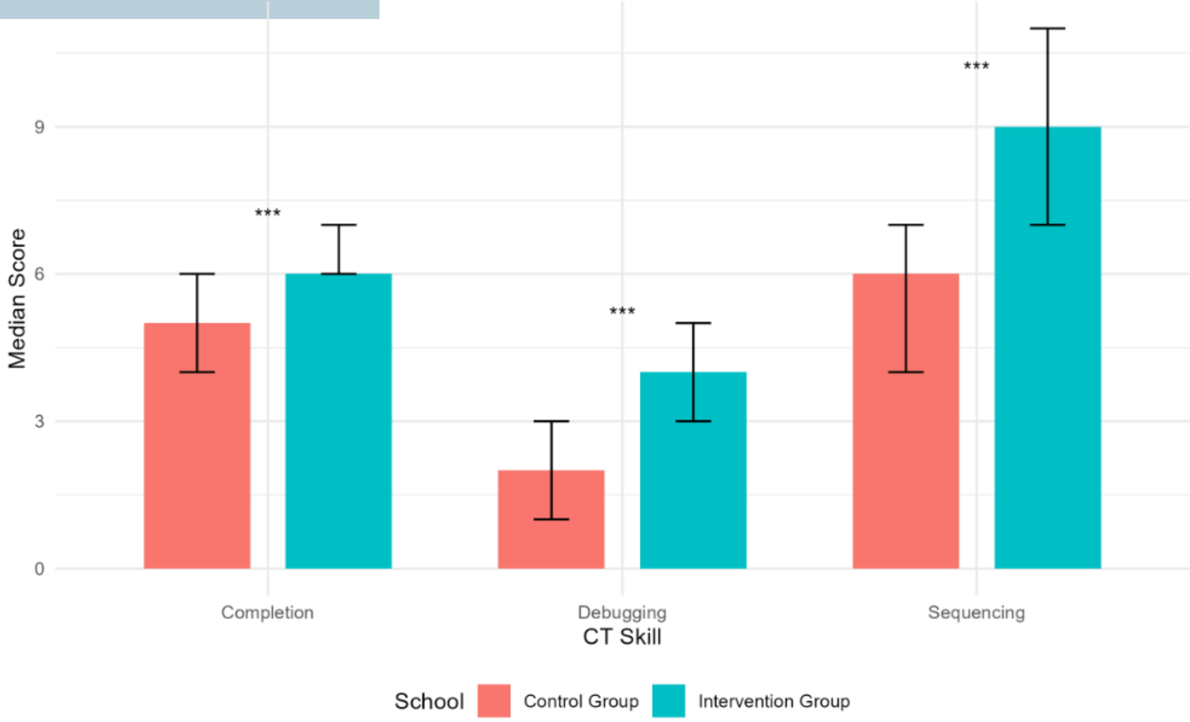


*Figure 2: Median scores by school for each Computational Thinking (CT) subskill, including significance levels of group differences.*

## Socio-Emotional Learning Scores

### Score calculation

Scores for the Social-Emotional Health Survey–Secondary (SEHS-S) were computed by averaging item responses at both the domain and subscale levels, following the structure of the original instrument.

The SEHS-S comprises four core domains of social-emotional health: Belief in Self, Belief in Others, Emotional Competence, and Engaged Living. Each domain includes three subscales, and each subscale is composed of three items. The calculation process was as follows:

- **Subscale scores** were calculated by averaging responses across the three items corresponding to each subscale.
- **Domain scores** were then calculated as the average of the three subscale scores within each domain.
- **Overall Covitality Score** was computed by averaging the four domain scores.

## Results

As shown in Appendix B, Shapiro-Wilk test indicated most of the socio-emotional learning scores and subscale scores did not follow a normal distribution ($p < .05$), especially in the intervention group. Although some subscales in the control group showed no significant deviation from normality, these were exceptions. Given the overall pattern of non-normality, non-parametric statistical methods were most suitable for group comparisons.

When examining the overall SEL score derived from the SEHS-S, a statistically significant difference was found between groups, favoring the intervention group ($W = 7615.5$, $p = 0.0087$), as shown in Figure 3. This suggests that students who participated in the intervention experienced higher overall levels of social-emotional well-being compared to their peers in the control group.

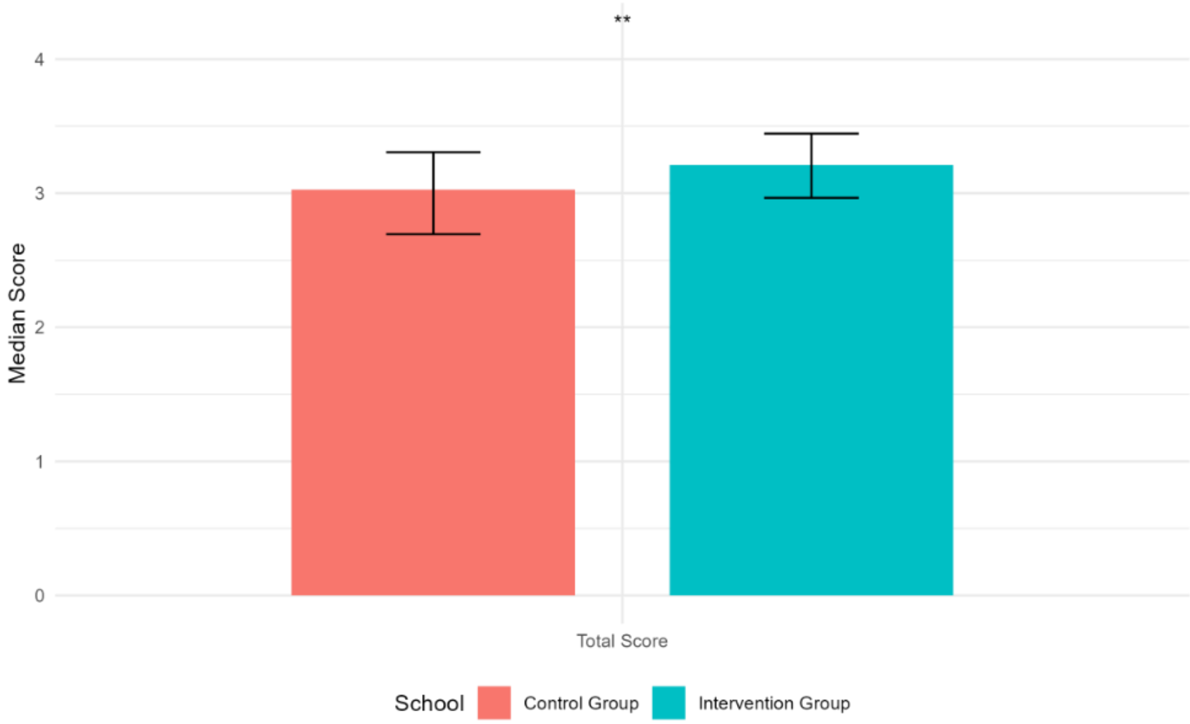


*Figure 3: Median total SEL scores by school, including significance level of group difference.*

At the domain level, the intervention group showed significantly higher scores than the control group for Belief in Others ($W = 7414$, $p = 0.027$), Emotional Competence ($W = 7494$, $p = 0.017$) and Engaged Living ($W = 7288$, $p = 0.050$), reflecting stronger perceived social support and emotional regulation, respectively (Figure 4).

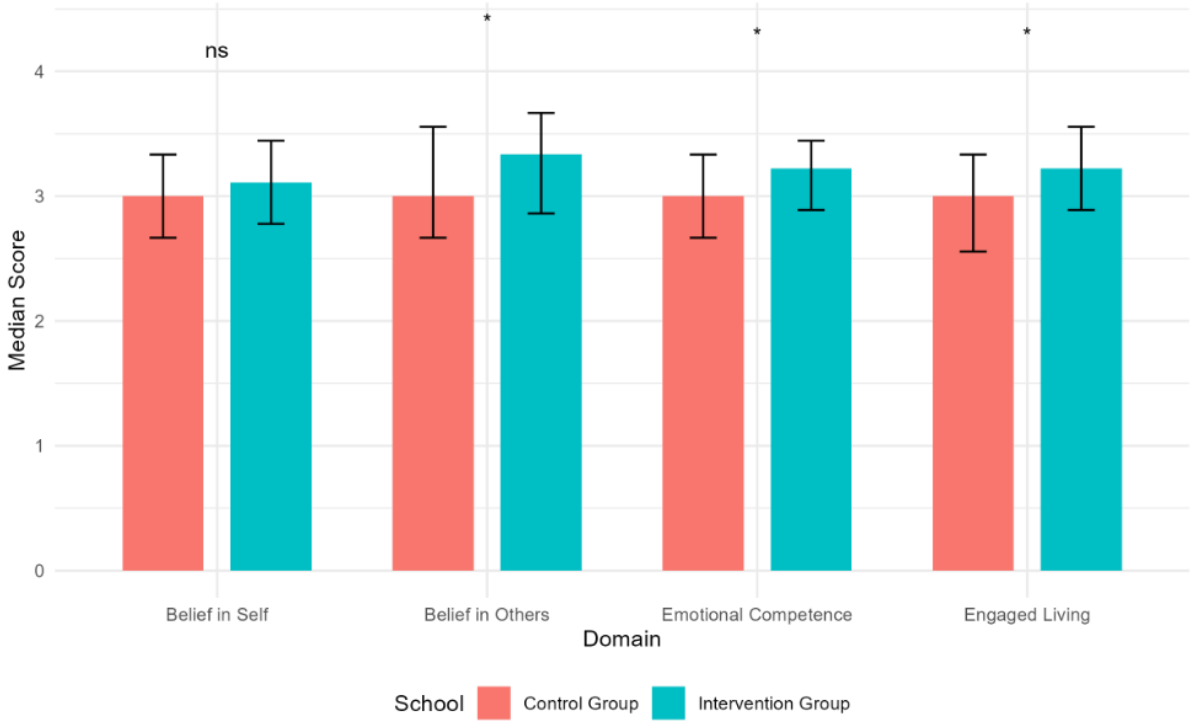


*Figure 4: Median scores by school for each SEL domain, including significance levels of group differences.*

At the subscale level, several statistically significant differences were observed in favor of the intervention group. Compared to controls, students in the intervention group reported higher scores in *Self-Efficacy* ($W$ = 8050, $p$ = 0.0004), *Peer Support* ($W$ = 7733, $p$ = 0.0034), *Empathy* ($W$ = 7847, $p$ = 0.0018), and *Optimism* ($W$ = 7646, $p$ = 0.0065), as shown in Figure 5.

Taken together, these results suggest that, compared to controls, the Arukay learning system is associated with enhanced social-emotional development, especially self-belief, interpersonal support, emotional skills, and future orientation. These findings offer evidence in relation to the second research question: Compared to standard practice, does Arukay improve SEL skills? The results indicate that the intervention is linked to higher levels of social-emotional learning outcomes among participating students than is typical for students receiving standard instructional practice.

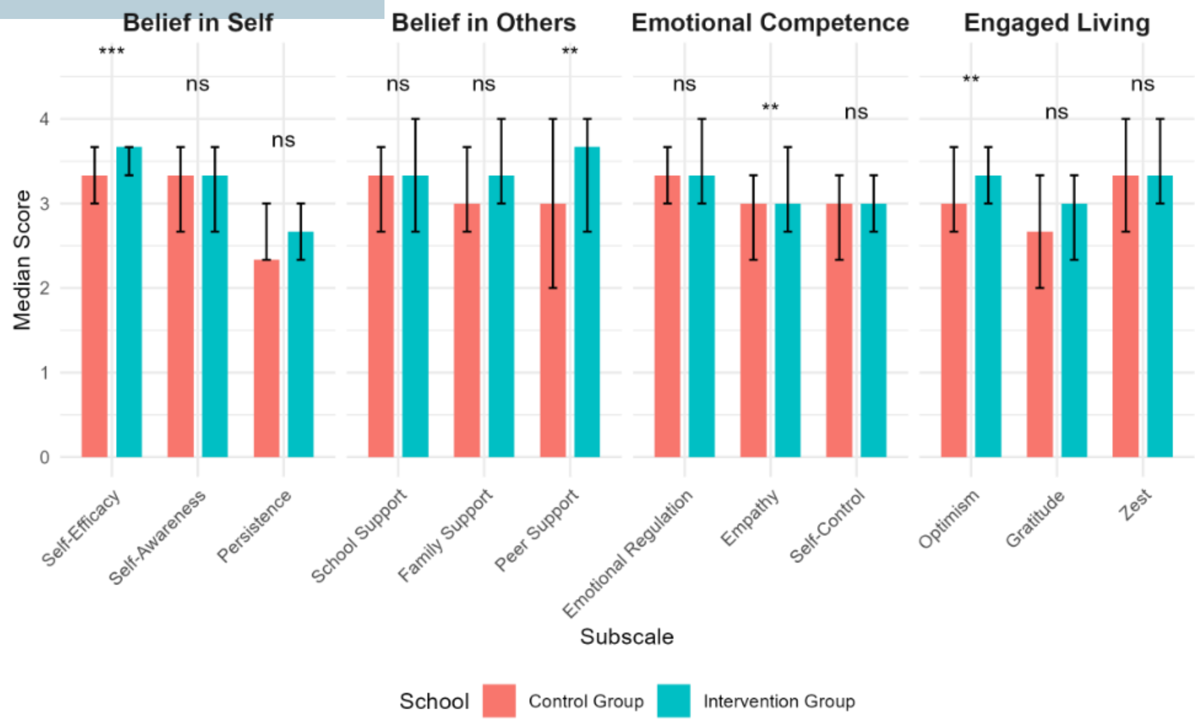


*Figure 5: Median scores by school for each SEL subskill, including significance levels of group differences.*

## Empathy-related Scores

### Score calculation

Scores for the Adapted Empathy in Design Scale (EMPA-D) instrument were calculated by averaging item responses across three conceptual dimensions: Emotional Interest and Perspective-Taking, Personal Experience, and Self-Awareness. Each dimension is comprised of 2 to 6 items, for a total of 11 items.

The calculation process was as follows:

- **Dimension scores** were calculated by computing the mean of the items corresponding to each of the three dimensions:
    - *Emotional Interest and Perspective-Taking* (6 items)
    - *Personal Experience* (2 items)
    - *Self-Awareness* (3 items)

- The **overall EMPA-D score** was derived by averaging responses to the 11 items across all dimensions.

Scores for the Basic Empathy Scale (BES) were computed by summing responses to a set of 20 items organized into four subdimensions: Positive-Cognitive Empathy, Negative-Cognitive Empathy, Positive-Affective Empathy, and Negative-Affective Empathy.

The calculation procedure was as follows:

- **Subdimension scores** were computed by summing item responses within each of the following categories:
  - *Positive-Cognitive Empathy* (5 items)
  - *Negative-Cognitive Empathy* (5 items)
  - *Positive-Affective Empathy* (5 items)
  - *Negative-Affective Empathy* (5 items)
- **General-Cognitive Empathy** and **General-Affective Empathy** scores were then derived by summing their respective subdimensions:
  - *General-Cognitive Empathy* = Positive-Cognitive + Negative-Cognitive
  - *General-Affective Empathy* = Positive-Affective + Negative-Affective
- The **overall BES score** was calculated as the sum of the *General-Cognitive Empathy* and *General-Affective Empathy* scores.

## Results

### *EMPA-D*

Shapiro-Wilk test results indicated that none of the EMPA-D scores or sub-scores met the assumption of normality in either the intervention or control group (see Appendix C). All p-values were below the .05 threshold. These results reinforce the appropriateness of using non-parametric statistical methods for analyzing group differences in empathy outcomes measured by the EMPA-D instrument.

Analysis of the overall EMPA-D score revealed a statistically significant difference favoring the intervention group ($W = 7997.5$, $p = 0.0007$), as shown in Figure 6. This suggests that participation in the Arukay intervention was associated with higher levels of empathy in design-related contexts compared to controls.

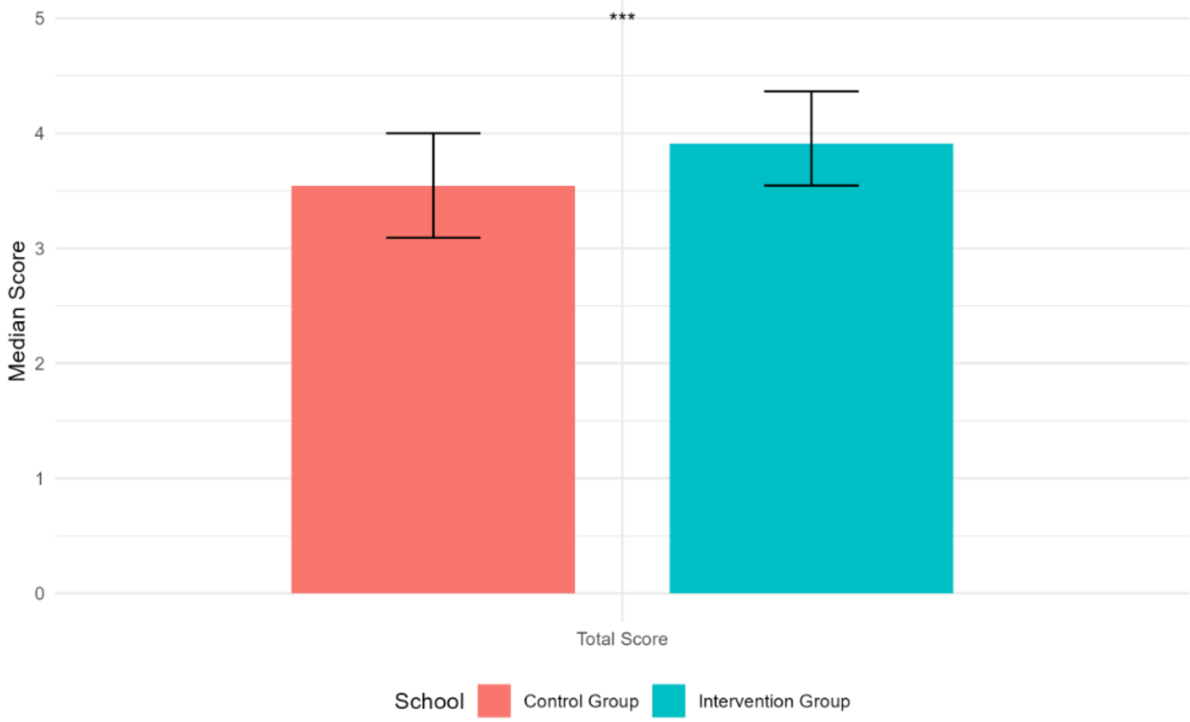


*Figure 6: Median total Empathy scores (EMPA-D) by school, including significance level of group difference.*

Statistically significant differences were also observed across all three dimensions of the EMPA-D, with the intervention group consistently scoring higher than the control group. The Emotional Interest and Perspective-Taking dimension showed a highly significant difference ($W = 8072$, $p < 0.001$), indicating a stronger capacity among intervention students to recognize and respond to others' emotions and viewpoints in design situations than controls. Significant differences were also observed in Personal Experience ($W = 7378$, $p = 0.029$), suggesting that students in the intervention group were more likely than controls to connect their own experiences to their empathetic engagement. Lastly, Self-Awareness scores were significantly higher in the intervention group than the control group ($W = 7860$, $p = 0.0016$), reflecting a deeper awareness of emotional and cognitive responses during design tasks (see Figure 7).

These findings suggest that, beyond supporting general social-emotional learning, the Arukay learning system fosters empathy-related skills that are particularly relevant to collaborative and user-centered problem solving to a greater extent than standard instructional practice. In relation to the third research question, compared to standard practice, does Arukay improve empathy-related skills, the results indicate that students who participated in the intervention demonstrated higher empathy scores across all assessed dimensions.

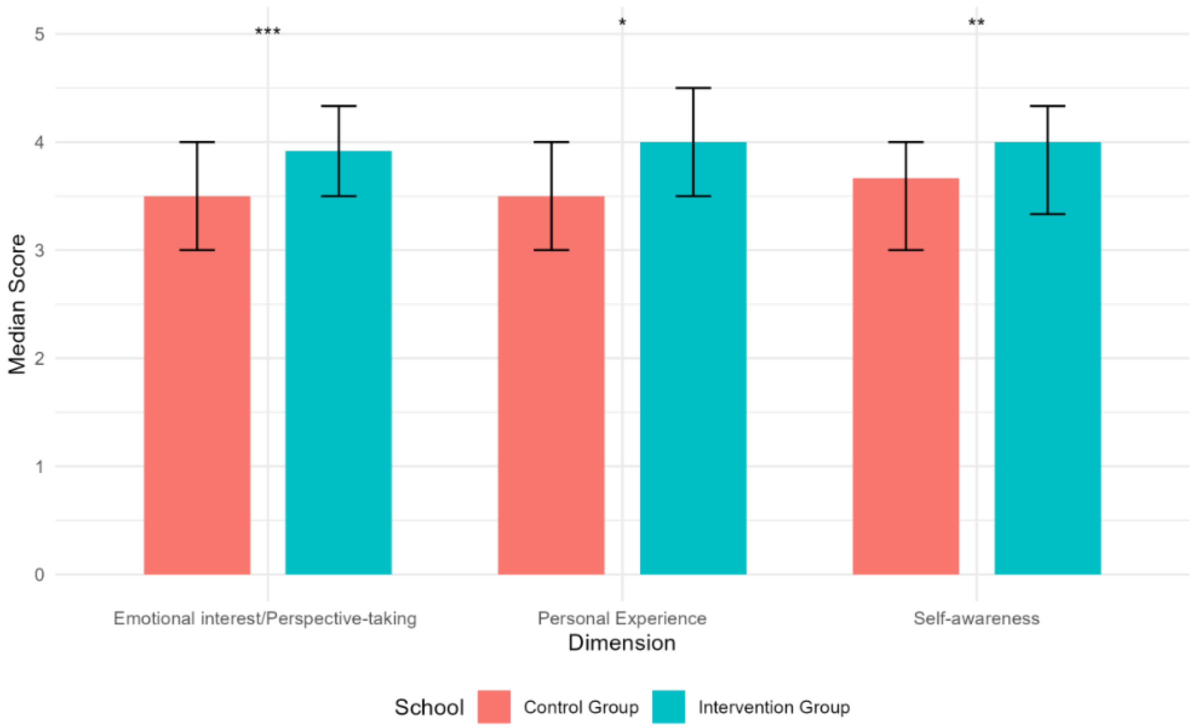


*Figure 7: Median scores by school for each Empathy subskill from the EMPA-D assessment, including significance levels of group differences.*

***BES***

Shapiro-Wilk test results revealed that the BES total score and all subdimension scores deviated significantly from a normal distribution in both the intervention and control groups (all p-values < .05; see Appendix D). These findings indicate that the distributional assumptions of parametric tests were not met, justifying the use of non-parametric statistical methods for evaluating group differences in empathy as measured by the BES.

Analysis of the overall BES score showed a statistically significant difference favoring the intervention group compared to controls (*W* = 7520.5, *p* = 0.0151). This suggests that students who received instruction through the Arukay learning system reported higher levels of general empathy compared to their peers in the control group (see Figure 8).

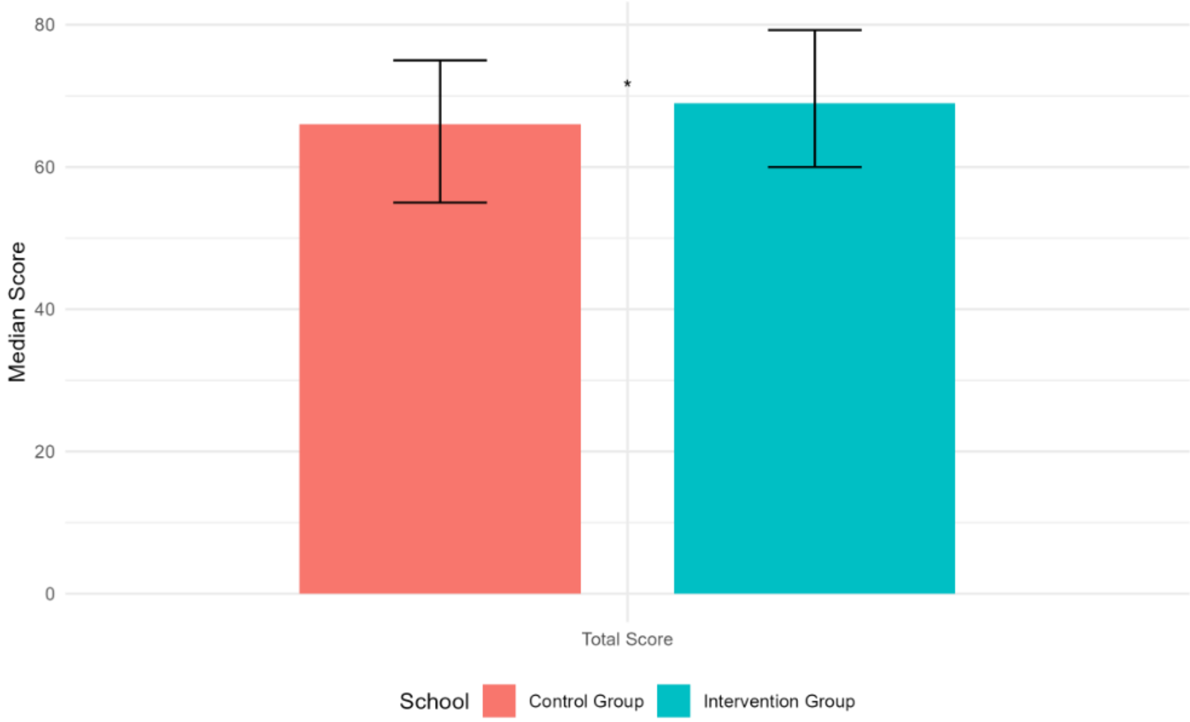


*Figure 8: Median total Empathy scores (BES) by school, including significance levels of group differences.*

At the composite level, a significant difference was observed in General-Cognitive Empathy (W = 7654, p = 0.0068), indicating stronger cognitive understanding of others' emotional states among students in the intervention group compared to controls (Figure 9). This pattern was further supported by results at the subdimension level: significant group differences were found in Negative-Cognitive Empathy (W = 7524, p = 0.0145), Positive-Cognitive Empathy (W = 7736, p = 0.0040), and Positive-Affective Empathy (W = 7448, p = 0.0221) (Figure 10). These findings suggest that the intervention group demonstrated enhanced ability to intellectually understand and positively resonate with the emotions of others than the control group.

Taken together, these results suggest that the Arukay learning system supports the development of empathy in adolescents, particularly in cognitive

domains and in the ability to engage with others' positive emotions to a greater extent than standard instructional practice. These findings provide evidence for the third research question with results from the BES indicating potential advantages of the Arukay learning system in key aspects of empathic
functioning.

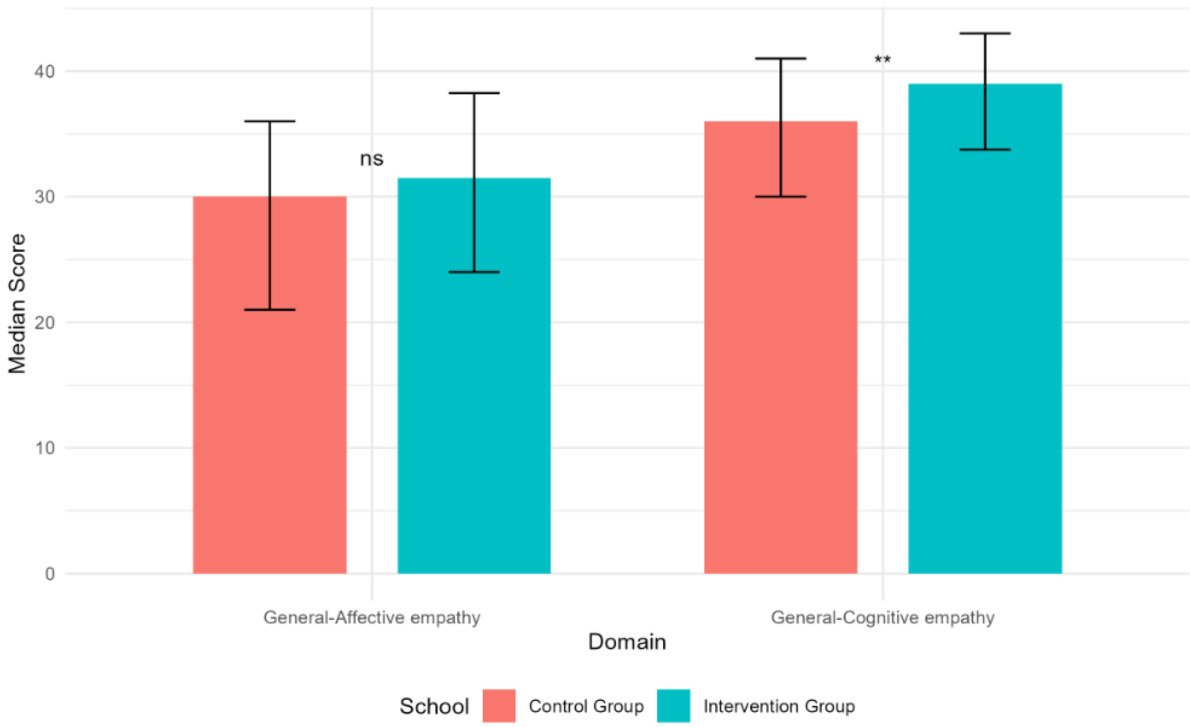


*Figure 9: Median scores by school for each Empathy (BES) dimension, including significance levels of group differences.*

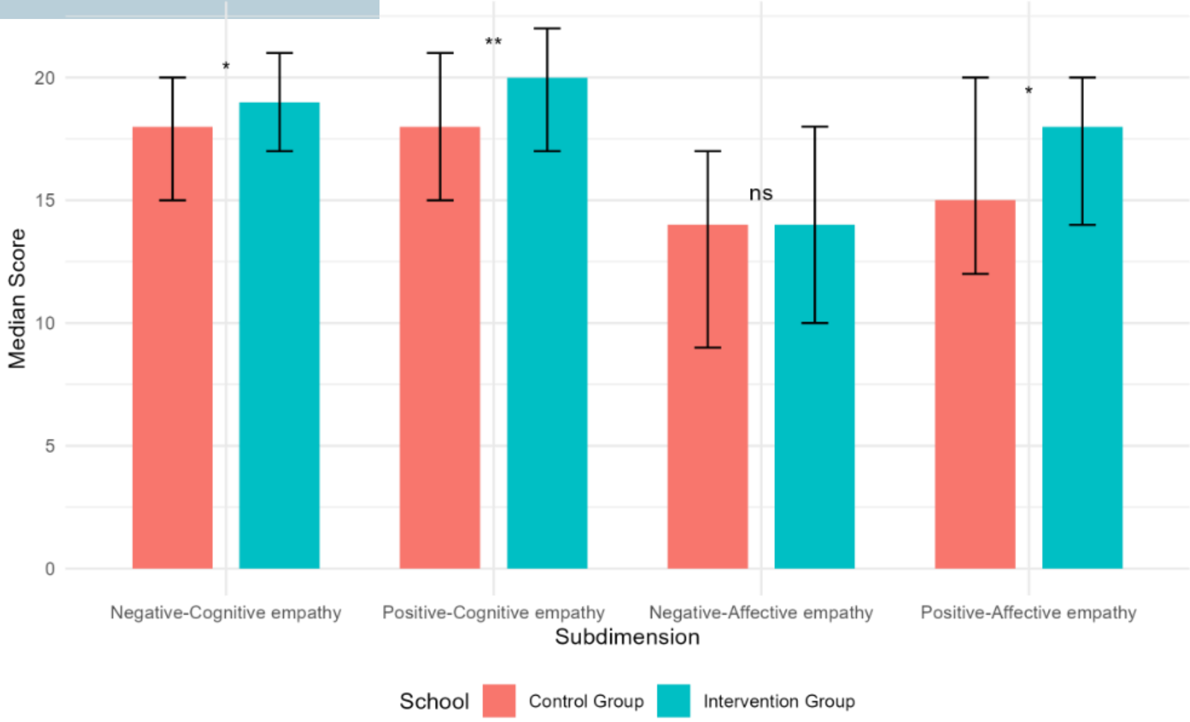


*Figure 10: Median scores by school for each Empathy (BES) sub-dimension, including significance levels of group differences.*

## Correlations between CT and SEL

The Spearman correlation matrices for both the intervention and control groups are available in Appendices E and F, respectively. A summary of the key findings per group is given below.

### Intervention group

For the intervention group, no statistically significant correlations were found between the total CT score and the total SEL score or any specific SEL domain.

Analyses at the SEL subskill level revealed that family support was positively associated with the total CT score ($\rho = 0.20$, $p = 0.019$). Further analysis of individual CT subskills revealed that sequencing skills were significantly and positively correlated with family support ($\rho = 0.23$, $p < 0.01$). Together, these results suggest that for the intervention group, students' perceptions of emotional and relational support from their families were associated with the development of computational thinking among students, and that students who feel supported at home are more proficient in organizing sequences of actions, a foundational aspect of computational thinking, than those who feel less supported by their families.

In addition, for the intervention group, debugging skills, a complex and problem-solving-oriented component of CT, were positively associated with persistence ($\rho = 0.18$, $p = 0.033$), as well as with emotional regulation ($\rho = 0.17$, $p = 0.049$). These findings highlight the potential importance of emotional competencies in supporting students' resilience and adaptability when facing challenges in programming and problem-solving tasks.

When comparing these correlation patterns with those of the control group, no statistically significant differences in the strength of the associations were observed. This may suggest that while certain social-emotional variables relate to CT performance, the intervention itself did not substantially enhance the strength of these relationships across the duration of this study.

**Control group**

In the control group, there were no significant correlations between the total CT score and overall SEL scores or domains, mirroring the findings from the intervention group. However, a notable exception was a positive correlation between persistence and the CT total score ($\rho = 0.24$, $p = 0.020$). This finding indicates that students who exhibit higher levels of persistence demonstrate stronger computational skills than those with lower levels of persistence, even in the absence of targeted instruction.

At the subskill level, no significant associations were found for sequencing and SEL scores, domains, or subdomains. In contrast, the completion subskill was positively associated with both the SEL total score and the Belief in Self domain ($\rho = 0.22$, $p = 0.031$; $\rho = 0.24$, $p = 0.020$, respectively). More specifically, completion was related to persistence ($\rho = 0.23$, $p = 0.028$), and Engaged Living ($\rho = 0.23$, $p = 0.030$), particularly with optimism and gratitude ($\rho = 0.22$, $p = 0.035$; $\rho = 0.21$, $p = 0.049$, respectively). These results suggest that students who can sustain effort and are optimistic and gracious perform better on tasks requiring the completion of multi-step processes than those whose effort wanes and have a pessimistic tendency lacking in gratitude.

The debugging subskill showed a broad pattern of association with SEL. It was positively correlated with the SEL total score and the Belief in Self domain ($\rho = 0.20$, $p = 0.055$; $\rho = 0.23$, $p = 0.030$), particularly with self-efficacy and persistence (ρ = 0.23, *p* = 0.025; ρ = 0.31, *p* = 0.002). These results emphasize the importance of intrapersonal strengths in supporting performance on tasks requiring error detection and correction, skills that are central to computational thinking.

**Summary**

Overall, whilst results did not yield strong direct associations between SEL and CT total scores, distinct patterns emerged across subskills and individual SEL dimensions in both groups. For both the intervention and control groups, persistence and self-efficacy correlated consistently with specific CT skills, especially completion and debugging, highlighting that intrapersonal factors and computational thinking are linked.

Notably, family support appeared influential only in the intervention group, suggesting that the family environment may moderate the relationship between social-emotional and computational competencies. These results underscore the value of integrating SEL components, particularly those that foster self-belief and sustained effort, into CT education, as they may serve as underlying drivers of student success, regardless of formal intervention. Further research is needed, however, to determine if a causal pathway between CT and SEL exists.

## Correlations between CT and Empathy

The Spearman correlation matrices for both the intervention and control groups are available in Appendices G and H, respectively. A summary of the key findings per group is given below.

**Intervention group**

For the intervention group, statistically significant negative correlations were observed between the total CT score and the total BES (ρ = -0.22, p = 0.010). This association was particularly evident with general-affective empathy (ρ = -0.27, p = 0.001) and one of its subcomponents, negative-affective empathy (ρ = -0.34, p < 0.001).

This pattern was mirrored at the subskill level, where sequencing was negatively correlated with the BES total score (ρ = -0.20, p = 0.017), general-affective empathy (ρ = -0.27, p = 0.001), and its subcomponent negative-affective (ρ = -0.33, p < 0.001). In addition, completion was negatively correlated with negative-affective empathy (ρ = -0.18, p = 0.039). No other correlations with the completion and debugging subskills and SEL subskills were significant.

These findings suggest a potential inverse relationship between CT performance, particularly sequencing, and certain dimensions of affective empathy, i.e., the ability to emotionally respond to and share the feelings of others. When comparing these correlation patterns with those of the control group, no statistically significant differences were observed in the strength of the associations. Further investigation is needed to determine if the pattern of negative correlations between CT skills and various empathy dimensions found here can be attributed to the Arukay learning system.

While the intervention group exhibited more consistent negative correlations between CT performance and certain dimensions of affective empathy, the absence of significant group differences suggests that these associations may not be directly attributable to the intervention. Instead, they could reflect broader cognitive-emotional patterns among adolescents. Further research is needed to clarify the nature of this relationship and to explore whether certain aspects of computational thinking are inherently linked to variations in empathic processing.

**Control group**

In the control group, none of the correlations between computational thinking and empathy were statistically significant.

**Summary**

Overall, these findings show a consistent pattern of negative associations between computational thinking skills, particularly sequencing, and affective components of empathy in the intervention group, with weaker and less consistent patterns observed within the control group. Although the intervention group exhibited more significant and systematic correlations than the control group, the lack of statistically significant differences in the strength of associations between groups warrants further investigation. Together, the results highlight a nuanced relationship between CT performance and empathy, warranting further investigation into how cognitive and emotional processes interact during learning in this age group, with or without targeted intervention.

# 10. Limitations

Analyses of group demographics revealed significant differences between the intervention and control groups in terms of age and socioeconomic stratum. The intervention group was of significantly lower age and higher socioeconomic stratum than the control group ($W = 4642$, $p < 0.001$; $W = 10722$, $p < 0.001$, respectively). Additionally, these two demographic variables were shown to be significantly correlated with some of the key outcome measures (CT, SEL, and empathy scores), suggesting possible confounding influences on results. Further exploration is required to understand how age and SES may influence the group differences and relationships between CT and SEL and empathy, as found above.

# 11. Recommendations for Future Research

This study has yielded key insights into the associations between CT and SEL, with a particular focus on empathy. In doing so, it has also identified several areas that merit further investigation:

- **Demographic influences.** Future research should explore in depth the role of age and socioeconomic status in shaping group differences and their relationship with CT, SEL, and empathy outcomes. A more detailed examination of these demographic variables could help to clarify their potential moderating or mediating effects.
- **Longitudinal and experimental designs.** While the current study provides promising evidence of the impact of the Arukay learning system, future research should adopt pre-post designs to better assess changes over time and investigate causal pathways. Extending the duration of the intervention could allow for a deeper development of student competencies and provide stronger evidence of the long-term effects of the intervention.

- **Broader student profiling.** Given the observed negative correlations between CT and certain dimensions of affective empathy, particularly in the intervention group, it is important to include additional variables in future studies that capture a broader profile of students' academic and social functioning. Variables such as general academic performance, teamwork skills, participation in extracurricular activities, and perspective taking may offer valuable context for interpreting patterns of empathy and computational thinking.

# 12. Practical implications for educators and policymakers

The findings of this study reinforce the growing consensus that SEL should be formally integrated across the school curriculum. Students who participated in the Arukay intervention showed enhanced competencies not only in CT but also in social-emotional domains such as emotional regulation, empathy, and belief in others. These results support the notion that SEL is not an isolated skill set, but one that is closely integrated with cognitive and problem-solving abilities.

For educators, this underscores the importance of implementing teacher training programs that embed SEL principles into everyday teaching practices, including subjects not traditionally associated with emotional development, such as technology or mathematics. Equipping teachers with strategies to foster students' emotional awareness, collaboration skills, and sense of purpose can enhance both academic and social outcomes.

For policymakers, these findings offer empirical support for recent legislative efforts in Colombia, including Laws 2414 (2024) and 2383 (2024), which mandate SEL integration into school curricula. The study highlights the urgency of translating these legal frameworks into scalable, evidence-based practices that reach diverse student populations. Interventions like Arukay, which support CT and SEL development through digital learning experiences, may offer promising models for implementation.

Moreover, this research suggests that computational thinking approaches may serve as a vehicle to support SEL development, especially when they incorporate elements of empathy, collaboration, and user-centered problem solving. Although further investigation is needed to clarify the mechanisms involved, such integrative approaches represent a promising pathway for addressing students' academic and emotional needs simultaneously.

# 13. Evidence Standards Alignment: ESSA Tier III and EduEvidence SILVER

This study aligns with Tier III ("Promising Evidence") under the 'Every Student Succeeds Act' (ESSA) and the SILVER level of efficacy according to the EduEvidence standards. Both frameworks recognize well-implemented correlational designs that demonstrate statistically significant associations between an intervention and relevant student outcomes.

The study employed a quasi-experimental design with a comparison group to examine the association between participation in the Arukay intervention and outcomes in CT, SEL, and empathy. Although not randomized, the design followed key elements of rigor: a well-defined intervention, clear comparison condition, and statistically appropriate methods for non-normally distributed data. This level of methodological rigor corresponds to Tier III under ESSA, and, given the positive and statistically significant results across multiple domains, aligns to EduEvidence SILVER criteria, which denotes promising efficacy based on correlational evidence from at least one study.

Statistically significant differences were observed between the intervention and control groups in several key domains. Specifically, students in the intervention group outperformed their peers in overall CT scores and in each CT subdimension ($p < .001$). Positive differences were also found in total SEL scores ($p < .001$) and in multiple subcomponents, including emotional competence and belief in others. Additionally, improvements were observed in empathy outcomes across both instruments used (EMPA-D and BES), particularly in the intervention group.

The Arukay learning system is a curriculum-based, scalable intervention that can be integrated into existing school schedules using standard technology infrastructure. The intervention was implemented in a public-school setting by a trained teacher, using digital tools and structured instructional guidance, making it feasible to replicate in similar educational environments. The learning modules used, Arukay Coding 5 and MakeCode Video Game Design, are accessible and adaptable, increasing the relevance of the intervention for diverse student populations.

The sample included students from two public schools in Bogotá, Colombia, with differing socioeconomic profiles: the intervention group was primarily composed of students from stratum 2, while the control group mainly included students from stratum 1. While each group was relatively homogeneous in terms of socioeconomic status, the inclusion of schools from distinct SES backgrounds broadens the relevance of the findings to a range of low-income educational settings within the Colombian public school system. While the sample was limited to two schools, the structure and delivery of the intervention suggest potential for generalizability, particularly in low- and middle-income urban public-school settings.

# 14. References

Candeias, A., Felix, A., Dumitrache, A., Almășan, B., Gencel, I. E., Zadworna, M., ... & Taskin, C. S. (2025). Enhancing socio-emotional learning and mental health through computational thinking: a cross-cultural analysis of the COMPUSEL programme. *Frontiers in Psychology*, *16*, 1480731.

Cheng, L., Wang, X., & Ritzhaupt, A. D. (2023). The effects of computational thinking integration in STEM on students' learning performance in K-12 education: A meta-analysis. *Journal of Educational Computing Research*, *61*(2), 416-443.

Collaborative for Academic, Social, and Emotional Learning (CASEL). (2025). *What is the CASEL framework?* https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/

Drouet, L., Bongard-Blanchy, K., & Lallemand, C. (2024). Development of the Empathy in Design Scale: Measuring Employees' Empathy Toward Users in Service Design. *Interacting with Computers*, 36(5), 313-334.

Ezeamuzie, N. O., & Leung, J. S. (2022). Computational thinking through an empirical lens: A systematic review of literature. *Journal of Educational Computing Research*, *60*(2), 481-511.

Kalelioglu, F., Gülbahar, Y., & Kukul, V. (2016). A framework for computational thinking based on a systematic research review. *Baltic Journal of Modern Computing*, *4*(3), 583.

Montuori, C., Gambarota, F., Altoé, G., & Arfé, B. (2024). The cognitive effects of computational thinking: A systematic review and meta-analytic study. *Computers & Education*, *210*, 104961.

Riaño-Hernández, D., Leenen, I., Ramírez-Conde, A., Atehortua-Rivera, P. A., & Piqueras, J. A. (2025). Validity evidence for assessing social-emotional psychological strengths in Colombian adolescents using the SEHS-S. PloS one, 20(2), e0314488.

Román-González, M. (2015). *Test de pensamiento computacional: Principios de diseño, validación de contenido y análisis de ítems* [Computational Thinking Test: Design guidelines, content validation and item analysis]. In M. Á. Murga Menoyo (Ed.), *Perspectivas y avances de la investigación* (pp. 279–302). UNED. Retrieved from https://www.researchgate.net/publication/288341872_Test_de_Pensamiento_Computacional_principios_de_diseno_validacion_de_contenido_y_analisis_de_items_Computational_Thinking_Test_design_guidelines_content_validation_and_item_analysis

Sakellariou, M. (2023). Using Computational Thinking To Strengthen Social-Emotional Learning In Education. In *The International Scientific Conference eLearning and Software for Education, 1*, 172-184.

Salas-Wright, C. P., Olate, R., & Vaughn, M. G. (2013). Assessing empathy in Salvadoran high-risk and gang-involved adolescents and young adults: A Spanish validation of the Basic Empathy Scale. *International Journal of Offender Therapy and Comparative Criminology*, 57(11), 1393-1416.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational research review*, *22*, 142-158.

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 103798.

# 15. Appendices

Appendix A. Normality Test Results (Shapiro-Wilk Test) for Computational Thinking Scores and Subscores

| Variable | Colegio La Bici (Intervention) | IED El Recuerdo (Control) |
|---|---|---|
| Total Score | W = 0.966, p = 0.0018<br>Non-normal | W = 0.978, p = 0.1183<br>Normal |
| Sequencing Subscore | W = 0.969, p = 0.0032<br>Non-normal | W = 0.975, p = 0.0728<br>Non-normal |
| Debugging Subscore | W = 0.871, p < 0.0001<br>Non-normal | W = 0.891, p < 0.0001<br>Non-normal |
| Completion Subscore | W = 0.925, p < 0.0001<br>Non-normal | W = 0.957, p = 0.0037<br>Non-normal |

Note. A p-value < .05 indicates a significant deviation from normality.

Appendix B. Normality Test Results (Shapiro-Wilk Test) for Socio-Emotional Learning Scores and Subscores

| Variable | Colegio La Bici (Intervention) | IED El Recuerdo (Control) |
|---|---|---|
| Total Score | W = 0.947, p = 0.01207<br>Non-normal | W = 0.987, p = 0.4736<br>Normal |
| Self-Efficacy Subscore | W = 0.871, p < 0.0001<br>Non-normal | W = 0.889, p < 0.001<br>Non-normal |
| Self-Awareness Subscore | W = 0.865, p < 0.0001<br>Non-normal | W = 0.882, p < 0.001<br>Non-normal |
| Persistence Subscore | W = 0.910, p = 0.0014<br>Non-normal | W = 0.935, p = 0.0029<br>Non-normal |
| Belief in Self Subscore | W = 0.929, p = 0.0054<br>Non-normal | W = 0.970, p = 0.1333<br>Normal |

| | | |
|---|---|---|
| School Support Subscore | W = 0.845, p < 0.0001<br>Non-normal | W = 0.860, p < 0.0001<br>Non-normal |
| Family Support Subscore | W = 0.839, p < 0.0001<br>Non-normal | W = 0.848, p < 0.0001<br>Non-normal |
| Peer Support Subscore | W = 0.840, p < 0.0001<br>Non-normal | W = 0.850, p < 0.0001<br>Non-normal |
| Belief in Others Subscore | W = 0.870, p = 0.0001<br>Non-normal | W = 0.950, p = 0.0128<br>Non-normal |
| Emotional Regulation Subscore | W = 0.840, p < 0.0001<br>Non-normal | W = 0.860, p < 0.0001<br>Non-normal |
| Empathy Subscore | W = 0.832, p < 0.0001<br>Non-normal | W = 0.912, p = 0.0046<br>Non-normal |
| Self-Control Subscore | W = 0.878, p = 0.0002<br>Non-normal | W = 0.900, p = 0.0050<br>Non-normal |
| Emotional Regulation Subscore | W = 0.925, p = 0.0160<br>Non-normal | W = 0.960, p = 0.0639<br>Normal |
| Optimism Subscore | W = 0.845, p < 0.0001<br>Non-normal | W = 0.915, p = 0.0002<br>Non-normal |
| Gratitude Subscore | W = 0.835, p < 0.0001<br>Non-normal | W = 0.920, p = 0.0006<br>Non-normal |
| Zest Subscore | W = 0.830, p < 0.0001<br>Non-normal | W = 0.900, p < 0.0001<br>Non-normal |
| Engaged Living Subscore | W = 0.929, p < 0.0001<br>Non-normal | W = 0.966, p = 0.0169<br>Normal |

Note. A p-value < .05 indicates a significant deviation from normality.

Appendix C. Normality Test Results (Shapiro-Wilk Test) for Empathy Scores and Subscores (EMPA-D assessment)

| Variable | Colegio La Bici (Intervention) | IED El Recuerdo (Control) |
| --- | --- | --- |
| Total Score | W = 0.965, p = 0.0015<br>Non-normal | W = 0.958, p = 0.0042<br>Non-normal |
| Emotional interest/<br>Perspective-taking | W = 0.964, p = 0.0012<br>Non-normal | W = 0.961, p = 0.0070<br>Non-normal |
| Personal Experience | W = 0.932, p < 0.0001<br>Non-normal | W = 0.930, p < 0.0001<br>Non-normal |
| Self-Awareness | W = 0.950, p < 0.0001<br>Non-normal | W = 0.939, p = 0.0003<br>Non-normal |

Note. A p-value < .05 indicates a significant deviation from normality.

Appendix D. Normality Test Results (Shapiro-Wilk Test) for Empathy Scores and Subscores (BES assessment)

| Variable | Colegio La Bici (Intervention) | IED El Recuerdo (Control) |
| --- | --- | --- |
| Total Score | W = 0.965, p = 0.0015<br>Non-normal | W = 0.958, p = 0.0042<br>Non-normal |
| Emotional interest/<br>Perspective-taking | W = 0.964, p = 0.0012<br>Non-normal | W = 0.961, p = 0.0070<br>Non-normal |
| Personal Experience | W = 0.932, p < 0.0001<br>Non-normal | W = 0.930, p < 0.0001<br>Non-normal |
| Self-Awareness | W = 0.950, p < 0.0001<br>Non-normal | W = 0.939, p = 0.0003<br>Non-normal |

Appendix E: Correlation Matrix between CT Scores and SEL Scores (Intervention Group)

Corr_table_intervention_CTvsSEL.xlsx

Appendix F: Correlation Matrix between CT Scores and SEL Scores (Control Group)

Corr_table_control_CTvsSEL.xlsx

Appendix G: Correlation Matrix between CT Scores and Empathy Scores (Intervention Group)

Corr_table_intervention_CTvsEmpathy.xlsx

Appendix H: Correlation Matrix between CT Scores and Empathy Scores (Control Group)

Corr_table_control_CTvsEmpathy.xlsx

## Acknowledgments

Arukay and ICEI acknowledge the funding from GOYN Bogotá and the Jacobs Foundation for this research.