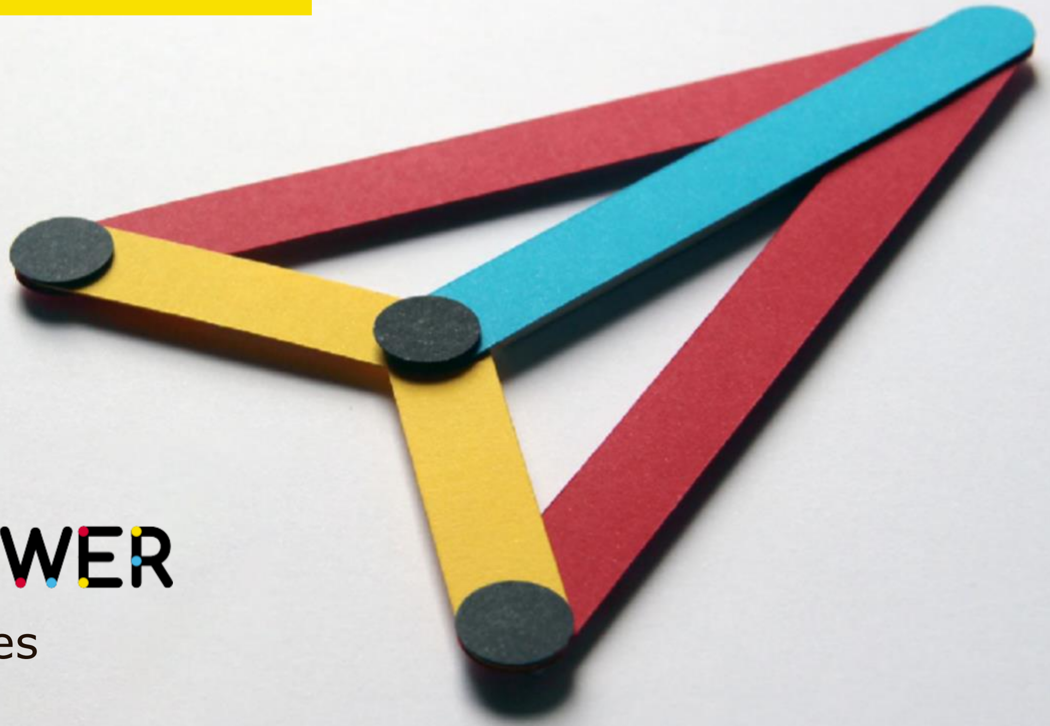
Deliverable Number

D6.4



EMPOWER

deliverables

Deliverable name	**Report with the results from RCT**
Type	**R - Document Report**
Dissemination level	**PU - Public**
Date	

Participants' description, instruments, setting, procedure, and the results of the study will be presented regarding the effectiveness of the intervention delivered through the technological platform compared to the standard intervention for children with NDDs.

Description

WP.6

Work Package. 7

UBB

Date	Version	Description	Authors
25.09.2025	1	First release of the document	Cristina Costescu, Lia Cozma, Oana Luiza Rebega, Carmen David, Adrian Roșan
30.09.2025	1.1	Update upon feedback	Gerardo Herrera, Diana Teixeira, Aristides Ferreira, Diana Teixeira, Sara Lopez, Paula Nunes Da Costa Ferreria,.

Contents

#1. Introduction *4*

1.1 Executive functions in NDD 4

1.2 Emotion Regulation in NDD 5

1.3 Assessing EF and ER using Technology 5

1.4 EMPOWER Platform – description 7

1.5 Objectives of our study 7

#2. Methodology *8*

2.1. Participants 8

2.2. Instruments 10

2.3. Procedure – Intervention Protocol 16

2.4. Results 19

#3. Conclusion *36*

3.1. Conclusion and discussions 36

3.2. Limitations 39

#4. References *40*

#1. Introduction

1.1 EXECUTIVE FUNCTIONS IN NDD

Research indicates that executive function deficits are prevalent across multiple neurodevelopmental disorders, though they manifest differently depending on the specific condition. Executive functions, including working memory, inhibition, and planning, are governed by the prefrontal cortex and play essential roles in goal-directed behavior and cognitive development (Sun & Buys, 2012). Studies demonstrate that flexibility is significantly impaired in both attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD). In contrast, specific learning disorders show less consistent executive function associations (Bausela Herreras et al., 2019). In developmental coordination disorder (DCD), executive function difficulties have been identified, though methodological issues complicate interpretation of findings (Leonard & Hill, 2015). Comparative research reveals that while both ASD and ADHD groups show executive function deficits compared to typical development, ASD children demonstrate greater socialization difficulties. In contrast, ADHD children show more impaired learning persistence. Importantly, different executive function components predict outcomes in each disorder, suggesting the need for differentiated intervention strategies (Roselló-Miranda et al., 2018).

EF significantly impact everyday functioning and academic outcomes in neurodivergent children. Research demonstrates that poor EF independently predicts school functioning, with interaction effects showing that children with high inattention symptoms and poor EF have a greater need for special education (Diamantopoulou et al., 2007). EF deficits are consistently associated with impaired social skills, behavioral problems, and reduced social participation across various NDDs (Tal-Saban et al., 2024). Specifically, spatial working memory and planning deficits contribute to school and peer dysfunctions in youth with ADHD, affecting academic performance, school interactions, and peer relationships (Chiang & Gau, 2014). Children with more severe EF profiles experience greater daily life impairments and increased parental stress, highlighting the cascading effects on family functioning (Pardo-Salamanca et al., 2024). These findings underscore the critical role of EF in determining functional outcomes across academic, social, and daily living domains in children with NDDs. Research has shown that impaired EF—including deficits in working memory, inhibition, and cognitive flexibility—limits academic success in children with Down syndrome, with unique patterns of influence compared to typically developing peers (Will et al., 2017). In autistic children, EF problems are also closely linked to everyday challenges, impacting their ability to plan, organize, and regulate behavior in daily situations, which in turn shapes their adaptive functioning and developmental trajectories (Kouklari et al., 2018; Lingala et al., 2021). Collectively, these findings underscore the critical role that EF plays in both school and day-to-day performance for children with NDDs, highlighting the need for targeted supports and interventions to foster academic and life skills in these populations.

1.2 EMOTION REGULATION IN NDD

Neurodivergent children demonstrate significant difficulties with emotion regulation that contribute to mental health problems. Research shows that cognitive control deficits in working memory, inhibition, and shifting predict increased emotion dysregulation, which fully mediates the relationship between cognitive challenges and both internalizing and externalizing symptoms (Tajik-Parvinchi et al., 2021). Across different neurodevelopmental conditions including autism spectrum disorder, and intellectual disability, individuals exhibit atypical emotion regulation strategy use, with maladaptive strategies like rumination and avoidance linked to elevated anxiety levels (Samson et al., 2022). These emotion regulation difficulties affect autonomy, social competence, and independent living skills (Hernández Lara et al., 2023). For example, Berkovits et al. (2014) found that children with developmental delays display greater emotion dysregulation than their typically developing counterparts, and this dysregulation is closely tied to long-term social problems and a weaker developmental trajectory of social skills. Poor ER abilities in NDDs have significant implications for everyday life and psychological health. Dysregulated emotion increases risk for both internalizing symptoms, such as anxiety and depression, and externalizing problems, such as aggression and impulsivity, regardless of whether a child belongs to a clinical or normative group (Cai et al., 2021). The maturation of ER relies on the development of interconnected prefrontal brain regions, and disruptions in these neurocognitive pathways in children with NDDs are linked to ongoing challenges with managing emotions, adapting to stress, and functioning at home, in school, or in community contexts (Martin & Ochsner, 2016). Parent emotion socialization processes further shape adolescents' ER and emotionality, underscoring the dynamic interplay between individual neurodevelopmental differences and the caregiving environment for neurodivergent youth (Breaux et al., 2021).

Interventions targeting emotion regulation, including cognitive-behavioral therapy and mindfulness training, show promise for improving outcomes (Mohd Hatta & Idris, 2025). However, technology-based supports for emotion regulation in this population remain limited and represent an emerging area needing further development (Hernández Lara et al., 2023).

1.3 ASSESSING EF AND ER USING TECHNOLOGY

Technological platforms are increasingly being developed to assess executive functions and emotion regulation in children with neurodevelopmental disorders (NDDs). Li et al. (2018) created a mobile gaming platform using social and nonsocial stimuli to assess executive function skills including cognitive flexibility, inhibitory control, and short-term memory in 65 children with and without ASD, successfully identifying diagnostic group differences and predicting IQ scores. Ramos-Galarza et al. (2020) developed a Python-based neuropsychological assessment platform for children with ADHD, measuring impulse control responses and reaction times, finding significant differences between ADHD and control groups. Cibrian et al. (2021) reviewed 36 studies examining emerging technologies across multiple platforms (computers, mobile devices, sensors, VR/AR, robots) for supporting self-regulation assessment and intervention in children with ADHD, noting promising results despite most technologies remaining in prototype phases.

Technology offers robust tools for both assessing and training executive functions (EF) and emotion regulation (ER) in neurodiverse children, enabling innovative, scalable, and engaging approaches to support these essential cognitive domains. Online cognitive testing platforms can identify unique cognitive profiles in children with neurodevelopmental disorders (NDD), such as ADHD and autism, providing more accessible alternatives to traditional in-person assessment while ensuring valid and reliable results (Hennessy et al., 2024). Broad, validated digital assessment tools—such as the Early Years Toolbox—further facilitate accurate early detection of EF challenges by measuring a range of executive, self-regulatory, and social development skills in children, helping to tailor supports and identify at-risk individuals efficiently (Howard & Melhuish, 2016). These advancements in technology-mediated assessment not only yield nuanced insights into children's strengths and weaknesses but also contribute data critical for intervention planning and progress monitoring.

Beyond assessment, technological solutions are increasingly being implemented for training and supporting EF and ER in neurodiverse populations. Systematic reviews demonstrate that digital interventions, gamified platforms, and virtual environments can strengthen children's executive functioning, foster self-regulation, and improve emotional well-being (Galitskaya et al., 2025). Technology-based programs designed specifically for individuals with developmental disabilities have proven effective in targeting ER, with tools ranging from biofeedback and mobile applications to immersive digital games that encourage skill practice in real-world and simulated contexts (Hernández Lara et al., 2023). Such innovations allow for tailored, adaptive training experiences that maximize engagement, provide immediate feedback, and can be used flexibly in homes, clinics, or schools, expanding access and long-term support for neurodiverse children as they navigate daily functional challenges.

While digital platforms for assessing and training executive function and emotion regulation in neurodiverse children offer accessibility and innovation, several limitations persist in their current form. Many computerized and mobile tools lack ecological validity, as they often evaluate well-structured tasks with clear goals and responses, which do not accurately reflect the complex, ambiguous challenges faced in real-life settings. As a result, performance on digital assessments may not translate directly to everyday functioning, especially in dynamic or socially nuanced scenarios. Furthermore, these platforms typically focus on a limited subset of cognitive domains, thus failing to capture the full spectrum of executive difficulties experienced by children with neurodevelopmental disorders.

Access and equity issues also present challenges, as effective use of technology-based assessments and interventions depends on reliable internet connectivity, device availability, and digital literacy among both families and educators. Moreover, the need for specialized supervision, data privacy, and adaptation to diverse cultural and linguistic backgrounds can complicate widespread implementation. Comorbidity and changing diagnostic criteria (such as evolving DSM editions) may further confound the interpretation of results, and long-term, longitudinal studies are still lacking to validate the predictive power of these tools for clinical outcomes. Thus, while technological platforms mark significant progress in neurodevelopmental assessment and training, their limitations highlight the need for ongoing research, greater real-world applicability, and improved accessibility for all children.

1.4 EMPOWER PLATFORM – DESCRIPTION

The Empower platform is designed to support education, psychological skill development, and research for neurodiverse children. It stands out through its strong use of gamification, transforming each skill into engaging serious games filled with diverse scenarios, avatars, rewards, and progressive challenges to sustain children's motivation and participation (Sailer et al., 2020). Games include tiered difficulty settings, which allow for individualized learning trajectories and ensure content accessibility for a wide range of abilities. The integration of smart devices like eye-trackers and smartwatches enables the platform to monitor both cognitive and physical activity during sessions, providing richer data for understanding each child's unique responses and needs. Teachers and caregivers can seamlessly monitor each student's progress, with the app delivering detailed performance analytics and algorithm-powered recommendations for personalized game selection and optimal challenge levels. See Deliverable 3.6 for a full description of the games.

1.5 OBJECTIVES OF OUR STUDY

This randomized controlled trial (RCT) is designed to rigorously evaluate the effectiveness of the EMPOWER intervention platform in supporting neurodiverse children. Central to the RCT is a comparative analysis of the EMPOWER platform against Treatment As Usual (TAU), specifically targeting changes in executive functioning, emotion regulation strategies, behavioral difficulties, and academic success. By examining outcomes across these domains, the study aims to determine whether children participating in the EMPOWER intervention exhibit measurable improvements relative to those receiving standard care. In addition to assessing intervention efficacy, the RCT also seeks to establish the validity and reliability of the EMPOWER platform as an innovative digital tool for psychological and educational support. These objectives are addressed through a set of focused research questions, providing a comprehensive framework for evaluating both the theoretical and practical impact of the platform on children's developmental trajectories.

This objective produces the following research questions (RQ):

RQ1. Is there a difference in executive functioning outcomes for children in the intervention group compared to the group receiving TAU at post-test compared to pre-test?

RQ2. How do children who experience the EMPOWER intervention differ in the use of adaptive emotional regulation strategies at post-test compared to pre-test, in comparison to children receiving TAU?

RQ3. What is the difference in behavior difficulties at post-test compared to pre-test for children who receive the EMPOWER intervention, as compared to those receiving TAU?

RQ4. Are there differences in academic success for children in the intervention group compared to the group receiving TAU at post-test compared to pre-test?

Another objective of this RCT is to evaluate the validity and reliability of the digital platform as an intervention tool, which produces one RQ more:

RQ5. Is the EMPOWER platform a valid and reliable intervention tool?

The above-mentioned objectives and research questions produce the following hypotheses (H):

H1. Children who experience the EMPOWER intervention will have better scores in executive functioning at the post-test compared to the pre-test than children with TAU.
H2. Children who experience the EMPOWER intervention will have more adaptive ER strategies at the post-test compared to pre-test than children with TAU.
H3. Children who experience the EMPOWER intervention will show less emotional and behavioral problems in the post-test compared to the pre-test than children with TAU.
H4. Children who experience the EMPOWER intervention will have better academic success and productivity in the post-test compared to the pre-test than children with TAU.
H5. The EMPOWER platform is a valid and reliable intervention tool.

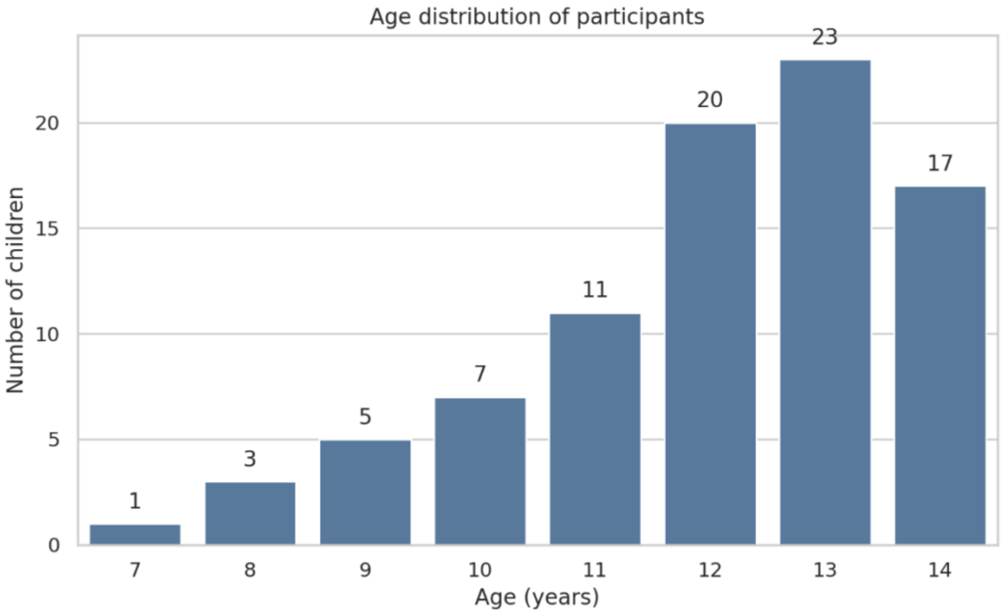
#2. Methodology

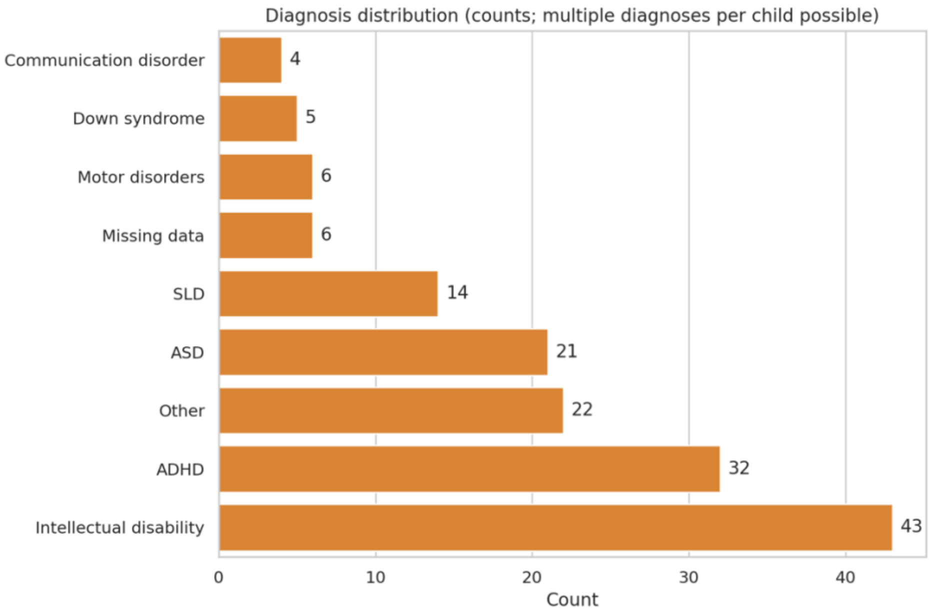
2.1. PARTICIPANTS

This study included 87 children diagnosed with neurodevelopmental disorders (NDD), recruited from Romania (n = 43) and Portugal (n = 44), to evaluate the effectiveness of the EMPOWER platform for training EFs and ER. The sample comprised 23 females and 64 males, with ages ranging from 7 to 14 years. Parent or legal guardian informed consent was obtained for all participants, and procedures adhered to ethical standards for research with minors and neurodiverse populations. The age distribution was skewed toward early to mid-adolescence: 81.6% of participants were aged 11–14, and 18.4% were aged 7–10, with the largest single-year groups at ages 12–14.

Inclusion criteria reflected the platform's requirements and the study aims. Eligible children were 7–14 years old and had a DSM-5-TR (APA, 2022) diagnosis of an NDD, including intellectual developmental disorder (including global developmental delay or not otherwise specified), autism spectrum disorder, attention-deficit/hyperactivity disorder (other specified or unspecified), specific learning disorder (reading, writing, or math), other specified or unspecified NDD, or any combination thereof. Consistent with the study's focus on executive functions and emotional self-regulation, classroom teachers observed difficulties in these domains. Participants were required to comprehend simple instructions presented orally and/or in writing and to have sufficient vision and motor abilities to engage with a screen and use common input devices (keyboard, mouse, or touch screen). For safety, children with epilepsy or implanted electronic devices (e.g., pacemakers, defibrillators, cochlear implants) were excluded. All included participants had adequate comprehension and basic familiarity with technology to use the platform.

Most children presented with more than one diagnosis, consistent with the high comorbidity typically observed in neurodevelopmental populations. The distribution of diagnoses included intellectual disability, ADHD, ASD, and SLD as the most frequent categories, followed by motor disorders, Down syndrome, and communication disorder; an “Other” category captured less common conditions (e.g., behavioral difficulties, developmental delay, GRIN2B, agenesis of the corpus callosum, global disorder, dyslexia, dysgraphia, Leigh syndrome, fetal alcohol syndrome, ataxic syndrome, and pyramidal syndrome). Six cases had missing diagnostic detail. Because comorbidities were common, diagnosis counts exceeded the number of unique participants.





2.2. INSTRUMENTS

To evaluate executive functioning and emotion regulation in school-aged children with neurodevelopmental disorders, we employed a multimethod, multi-informant battery that combines standardized rating scales completed by parents and teachers with direct, performance-based tasks. Core executive processes—working memory, planning, inhibition, attention, and cognitive flexibility—were indexed using the Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008), the Executive Skills Checklist (ESC; Diamond, 2016; Diamond & Ling, 2016), and the REMEX battery of direct tasks (Costescu, David, Roșan, Cozma, & Pădure, under review). Emotion-related competencies were assessed using the Academic Emotion Regulation Questionnaire (AERQ; Burić, Sorić, & Penezić, 2016) and the Assessment of Children's Emotion Skills (ACES; Schultz, Izard, & Bear, 2004), capturing both regulation strategies and emotion knowledge/recognition. To support validation analyses and establish convergent and criterion-related evidence for the EMPOWER platform, we included two external criteria: the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) and the Academic Performance Rating Scale (APRS; DuPaul, Rapport, & Perriello, 1991). Across measures, internal consistency in the current sample ranged from acceptable to excellent, providing a reliable foundation for examining executive and emotional functioning in this 7–14-year-old neurodiverse cohort.

Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008)

The CHEXI is a 24-item rating scale designed for parents and teachers to assess core executive function domains in children. The instrument targets working memory, planning, inhibition, and regulation, providing a comprehensive index of everyday executive functioning in naturalistic settings. The developers report acceptable test–retest reliability over a 2–4-week interval, with coefficients of 0.89 for the total score, 0.86 for the inhibition subscale, and 0.75 for the working memory subscale. In the current sample, internal consistency was excellent (Cronbach's alpha = 0.96), supporting the reliability of the total score for research and clinical applications in neurodiverse populations.

Executive Skills Checklist (ESC; Diamond, 2016; Diamond & Ling, 2016)

The ESC is a behaviorally anchored checklist that indexes executive skills in school-aged children across day-to-day contexts. It includes an attention scale and a cognitive flexibility subscale relevant to goal maintenance, distractor resistance, and set shifting. In the current sample, internal consistency was excellent for both focal scores, with Cronbach's alpha of 0.90 for the attention scale and 0.92 for the cognitive flexibility subscale, indicating robust reliability for use as proximal indicators of executive control in the target population.

REMEX – Assessing Executive Functions in Children: Direct Tasks (Costescu, David, Roșan, Cozma, & Pădure, under review)

REMEX is a performance-based battery of direct executive function tasks administered to children. It focuses on the component processes of executive control through structured tasks that elicit working memory, attentional control, inhibitory control, and self-regulatory processes. As a direct assessment, REMEX complements informant reports by indexing executive functioning under standardized conditions. In this study, REMEX included, among other tasks, a visual attention task adapted from NEPSY and a delay-of-gratification task adapted from prior experimental paradigms (see below), providing convergent process-level measures of attention and self-control.

Visual Attention Task (adapted from NEPSY; Korkman, Kirk, & Kemp, 2007)

This task assesses selective and sustained attention via child-friendly, time-limited visual search requirements. Children identify target stimuli among distractors under standardized instructions. Performance-based indices are sensitive to attentional control and processing efficiency and are widely used in neurodevelopmental assessments. The adaptation employed here preserved the core structure of NEPSY's visual attention paradigm while aligning with the study's lab-based procedures.

Delay of Gratification Task (adapted from Kochanska et al., 1996)

This task probes self-regulatory capacity and inhibitory control by requiring children to delay an immediate, smaller reward to obtain a larger, delayed reward. Performance reflects impulse control and strategic regulation in a structured context. The adapted procedure followed established protocols while maintaining standardized instructions and timing to ensure comparability across participants.

Academic Emotion Regulation Questionnaire (AERQ; Burić, Sorić, & Penezić, 2016)

The AERQ is a 16-item self-report measure targeting strategies and tendencies for regulating emotions in academic contexts. It captures how children manage emotional states that arise during learning and performance. In the current sample, internal consistency was acceptable (Cronbach's alpha = 0.88), supporting its use for research on emotion regulation in school-related settings among neurodiverse youth.

Assessment of Children's Emotion Skills (ACES; Schultz, Izard, & Bear, 2004)

The ACES assesses foundational emotional competencies across three subtests: recognition of facial expressions, recognition of emotion from behavioral descriptions, and emotion knowledge derived from situational vignettes. The instrument comprises 56 items and evaluates children's ability to recognize emotional arousal and identify emotion–context contingencies. Reliability reported in the literature is mixed across studies (Vergara et al., 2022); however, most findings indicate Cronbach's alpha values acceptable for research applications, supporting its use to index emotional understanding relevant to classroom functioning and social communication.

Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)

The SDQ is a widely used 25-item behavioral screening questionnaire covering emotional symptoms, conduct problems, hyperactivity–inattention, peer relationship problems, and prosocial behavior. It provides a balanced profile of strengths and difficulties relevant to mental health and school functioning. In the current sample, internal consistency for the total score was acceptable (Cronbach's alpha = 0.72), meeting conventional thresholds for research use.

Academic Performance Rating Scale (APRS; DuPaul, Rapport, & Perriello, 1991)

The APRS is a 19-item teacher rating scale that evaluates academic success and classroom functioning, including task completion, productivity, and learning behaviors. In the current sample, the APRS demonstrated excellent internal consistency (Cronbach's alpha = 0.91), supporting its utility as a criterion indicator of academic functioning for validation analyses.

Administration overview and rationale

The battery integrates multi-informant questionnaires (parents, teachers, and self-report where applicable) with direct, performance-based tasks to capture both everyday executive behaviors and underlying control processes. This multi-method approach enhances construct coverage, addresses shared-method variance, and supports convergent validity testing for the EMPOWER platform. Reliability indices observed in the current sample were generally acceptable to excellent, providing confidence in the stability of the measures for the target neurodiverse population.

Game-based measurements

This game-based measurement matrix operationalizes core executive functions and emotion regulation into concrete, child-friendly performance indicators collected across multiple mini-games. Attention is indexed in Mushroom Hunters through accuracy on target-present and target-absent trials, error types (misses and distractor hits), and mean response time, capturing sustained attention and vigilance. Working memory is assessed in WorM via rule-based sorting accuracy, order-sensitive recall, longest correctly recalled sequence, and response latencies, reflecting maintenance, updating, and serial recall. Inhibition is measured in ReStroop by contrasting congruent and incongruent trial accuracy, commission and omission errors, the percentage of unanswered trials, and condition-specific response times, indexing interference control and response inhibition. Emotion processes are captured through in-game self-reports and choices: the Wheel, Cornhole, and Egg Pyramid tasks record children's current emotion, perceived intensity, understanding of emotion causes, and the selection of adaptive versus maladaptive regulation strategies; Cornhole additionally logs motor-accuracy throws alongside emotion labeling and matching of facial expressions. Collectively, the matrix yields count-based outcomes (corrects, errors), proportions (e.g., unanswered trials), and latency metrics that together provide converging, task-diverse evidence for executive control and emotion regulation in an engaging, ecologically valid format (See Table 1).

Table 1. Detailed measurement matrix for EMPOWER game tasks

Ability	*Measure*	*What it means*
		GAME: ***Mushroom Hunters – Attention game***
Ability to maintain attention throughout the task.	Correct Present	Correct responses when the mushroom is present (i.e., hits).
	Correct Absent	Correct responses when the mushroom is not present (i.e., does not hit distractors)
	Error Present	Incorrect responses when the mushroom is present (i.e., misses).
	Error Absent	Incorrect responses when the mushroom is not present (i.e., hits distractors)
	Mean Response Time	Mean of the response time
		GAME: ***WorM – Working memory game***
Ability to sort items based on two criteria (peppers with worm or peppers without worm)	Correct Pepper Sort	Number of peppers correctly sorted into the worm/no-worm baskets
	Incorrect Pepper Sort	Number of peppers incorrectly sorted into the worm/no-worm baskets
Ability to retain a set of items (peppers) in memory, considering their position and order, so they can be recalled after a short period of time	Correct Order Recall	Total number of peppers correctly recalled in the same order as they were originally presented
	Wrong order Correct recall	Total number of peppers correctly recalled but not in the order in which they were originally presented
	Longest Sequence	Highest number of peppers correctly selected in the correct order
	Mean Response Time	Mean of the response time
		GAME: ReStroop – Inhibition game
Ability to control impulses and resist distractions or automatic responses (in this case, sort by shape, not by color)	Correct Congruent	Total correct answers in congruent trials
	Correct Incongruent	Total correct answers in incongruent trials
	Congruent Errors	Total incorrect answers in congruent trials
	Incongruent Errors	Total incorrect answers in incongruent trials
	Commission Errors	Errors resulting from incorrect responses, excluding omissions (lack of response)
	Omission Errors	Errors due to lack of response or responses made after the time expired (during the trash-sorting stage of the game)
	Percent Trials No Answers	Percentage of trash items left unclassified because the child did not select them
	Mean Response Time	Mean of the response time
	Mean Response Time Congruent	Mean response time for congruent trials

	Mean Response Time Incongruent	Mean response time for incongruent trials
		GAME: ReFlex – Cognitive Flexibility game
Ability to switch between different rules based on changing game demands	Total Trials	Total number of trials displayed
	Correct Responses	Total correct responses
	Errors	Total number of errors
	Category Correct Sort	Total number of correctly sorted categories: color, shape, and number
	Perseverative Errors	Total number of errors resulting from participants persisting with a previous sorting rule or strategy after being told it was incorrect
	Non-perseverative Errors	Total number of errors resulting from mistakes not caused by using an outdated rule, such as guessing or misapplying the new rule
	Failure Maintain Set	Number of times a participant fails to sort images correctly after sorting at least five in a row. If any of images 6–10 are sorted incorrectly, it counts as a failure to maintain the set
	Trials For First Category	Number of trials needed to correctly sort the first category
	Mean Response Time	Mean of the response time
		GAME: BeeHold – Delay of Gratification game
Ability to resist immediate rewards in order to achieve a larger reward later	Able to wait Trial 1	True/false the child was able to wait for trial 1
	Able to wait Trial 2	True/false the child was able to wait for trial 2
	Able to wait Trial 3	True/false the child was able to wait for trial 3
	Wait Time Trial 1	Time that the child was able to wait for trial 1
	Wait Time Trial 2	Time that the child was able to wait for trial 2
	Wait Time Trial 3	Time that the child was able to wait for trial 3
		GAME: EmotiFest – Emotions games
Ability to recognize and match facial emotional expressions conveying the same emotion	Wheel Naming_Correct Match	Number of correct matches of emotion expression pictures in the wheel game
Ability to recognize one's own emotions	Wheel Emotion	Emotion selected by the child during the wheel game
Ability to recognize the intensity of one's own emotions	Wheel Intensity	Quantity of potion selected by the child in the wheel game, indicating emotion intensity (range: 1–3)
Ability to understand the	Wheel Understanding	Emotion understanding option chosen by the child in the wheel game

reasons behind one's own emotions		
Ability to regulate one's own emotions using adaptive strategies	Wheel Regulation_option adaptive	Adaptive emotion regulation strategies options selected in the wheel game
Ability to regulate one's own emotions using maladaptive strategies	Wheel Regulation_option maladaptive	Maladaptive emotion regulation strategies options selected in the wheel game
Ability to recognize and match facial emotional expressions conveying the same emotion	Cornhole Naming_correct match	Number of correct matches of emotion expression pictures in the cornhole game
Abilities involving fine and gross motor skills, accuracy, and focus	Cornhole Naming_correct throw	Number of correct throws (direction and hit the hole) in the cornhole game
Ability to recognize one's own emotions	Cornhole Emotion	Emotion selected by the child during the cornhole game
Ability to recognize the intensity of one's own emotions	Cornhole Intensity	Quantity of potion selected by the child in the cornhole game, indicating emotion intensity (range: 1–3)
Ability to understand the reasons behind one's own emotions	Cornhole Understanding	Emotion understanding option chosen by the child in the cornhole game
Ability to regulate one's own emotions using adaptive strategies	Cornhole Regulation_option adaptative	Adaptive emotion regulation strategies options selected in the cornhole game
Ability to regulate one's own emotions using maladaptive strategies	Cornhole Regulation_option maladaptative	Maladaptive emotion regulation strategies options selected in the cornhole game
Ability to recognize one's own emotions	Egg Emotion	Emotion selected by the child during the egg pyramid game

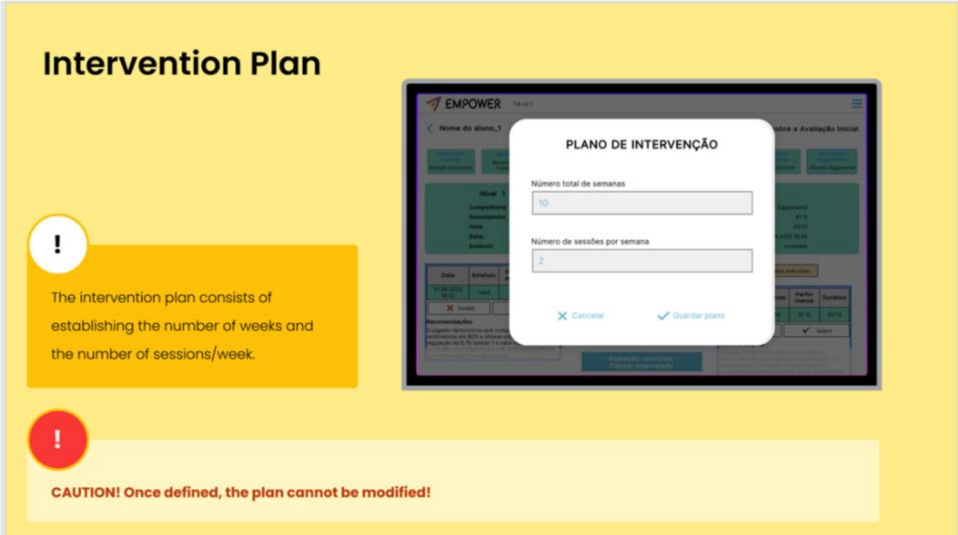
Ability to recognize the intensity of one's own emotions	Egg Intensity	Quantity of potion selected by the child in the egg pyramid game, indicating emotion intensity (range: 1–3)
Ability to understand the reasons behind one's own emotions	Egg Understanding	Emotion understanding option chosen by the child in the egg pyramid game
Ability to regulate one's own emotions using adaptive strategies	Egg Regulation_option adaptative	Adaptive emotion regulation strategies options selected in the egg pyramid game
Ability to regulate one's own emotions using maladaptive strategies	Egg Regulation_option maladaptative	Maladaptive emotion regulation strategies options selected in the egg pyramid game

2.3. PROCEDURE – INTERVENTION PROTOCOL

The intervention using the EMPOWER platform consisted of 20 sessions, each lasting 30 minutes, conducted over a 10-week period at a frequency of two sessions per week. During each session, participants engaged with three different games, each targeting specific executive functions or emotional regulation skills. The games included: *Mushroom Hunters* (sustained attention), *WorM* (working memory), *ReStroop* (inhibition), *ReFlex* (cognitive flexibility), *BeeHold* (delay of gratification), and *Emotifest* – three emotion-focused games (*Wheel*, *Cornhole*, *Egg Pyramid*).

A key component of the intervention was the Teacher's Application, accessible via tablet, which allowed the teacher to monitor each student's performance in the games played. The application provided recommendations regarding the games and levels the child should play during each session, based on previously recorded performance. These algorithm-based suggestions could be accepted or modified by the teacher, who also considered observational data gathered during gameplay.
At the outset, each student underwent an initial assessment during which they played all the games at least at the first level. Advancement to higher levels was dependent on reaching a minimum performance threshold of 70%.

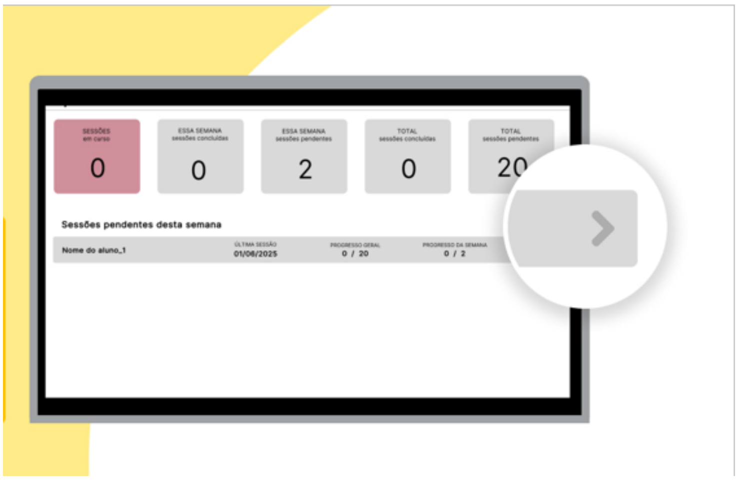
Before the intervention commenced, an individual intervention plan was established for each student. In this study, a standardized intervention plan was set, comprising 20 sessions spread over 10 weeks, with two sessions per week.



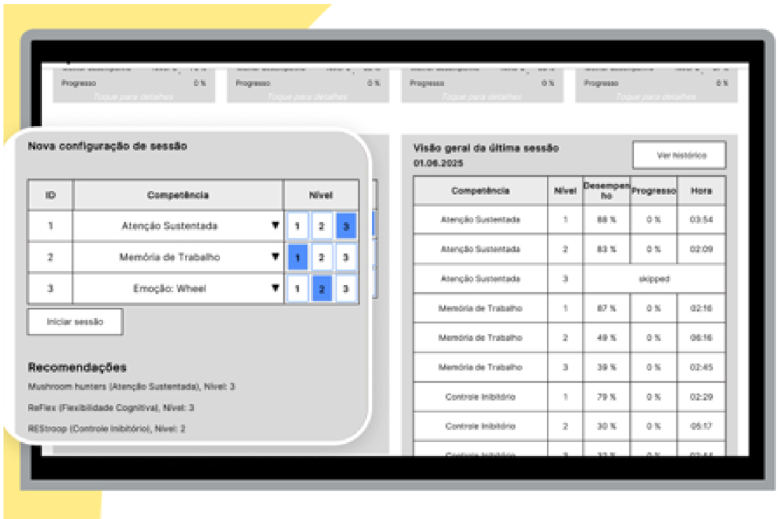
During the first session, the teacher could review the student's performance from the initial assessment in the Teacher Application, alongside recommendations generated by the application's algorithms regarding which games and levels the student should play, based on identified skills needing improvement. For instance, if a student achieved only 30% accuracy on the first level of the WorM game but reached over 70% accuracy even on the second level of ReStroop, the algorithm would likely recommend the WorM game for intervention, while excluding ReStroop from the first session plan.

Based on these insights, the teacher can choose to follow the algorithm's suggestions or make adjustments, taking into account external factors not captured by the system. For example, even if the student performed well on ReStroop, the teacher might still decide to include it in the session if they suspect the student was guessing answers and had not truly mastered the skill. Conversely, the teacher might opt not to include WorM in the session if they believe the student is currently unable to engage effectively in this specific game—due to fatigue, distraction, or a need to strengthen foundational skills such as attention. In such cases, they might prioritize an alternative game like Mushroom Hunters instead.

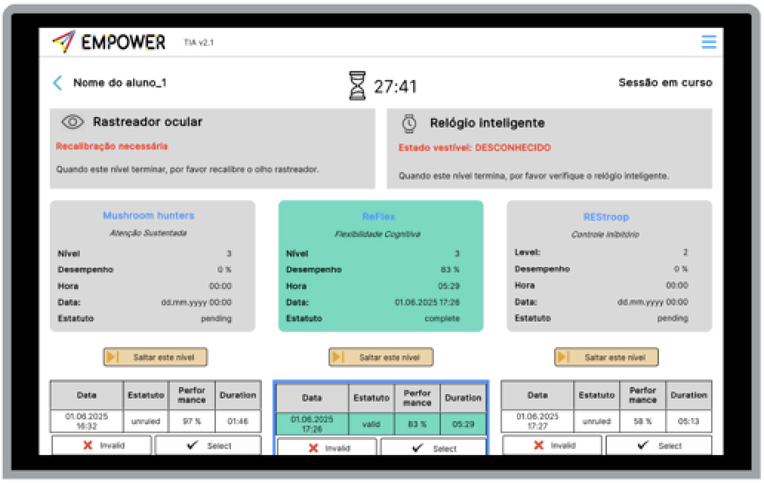
At the beginning of each session, the teacher logged into the EMPOWER game on a laptop and accessed the Teacher Application on the tablet, selecting the student for the current intervention session. The screen displayed the student's weekly schedule, including the current session, the number of sessions planned and completed for that week, as well as the total sessions scheduled and completed overall.



The next screen allowed the teacher to plan the upcoming session (comprising three games) and view statistical information on the student's progress for each game and details of the most recent session. To set up a new session, the application suggested a set of three games and level combinations. Each session must include at least one game targeting emotional competencies. Therefore, if two of the selected games are *Mushroom Hunters* and *ReFlex*, the third must be chosen from among the emotion-focused games: *Wheel*, *Cornhole*, or *Pyramid*, which are designed to train emotion recognition, understanding, and regulation.



Subsequently, the session monitoring screen displayed the three games and corresponding levels to be played. As the student progressed through the games on the laptop, the teacher could observe their activity and performance via the application, confirm correct completion of tasks, or skip tasks if the student was unable to complete them. The session concluded once all planned games were confirmed.



2.4. RESULTS

Executive functions

Using two-tailed t-tests, we compared groups on all pre-test measures. Descriptively, group means were comparable across measures, and no test reached significance (See Table 2). For example, Working Memory was slightly higher in Experimental group than Control group (M = 32.07, SD = 8.12 vs. M = 29.07, SD = 6.51), but the difference was not significant, $t(79.7) = 1.87$, $p = .065$. Similarly, Planning, $t(81.5) = 1.15$, $p = .253$; Regulation, $t(81.7) = 1.23$, $p = .223$; Inhibition, $t(81.4) = 0.23$, $p = .815$; Total Working Memory, $t(80.0) = 1.68$, $p = .096$; Total Inhibition, $t(81.5) = 0.49$, $p = .624$; Attention, $t(80.5) = 0.55$, $p = .585$; and Flexibility, $t(81.9) = 0.94$, $p = .350$, all failed to reach the $\alpha = .05$ threshold. Collectively, these results indicate no reliable pre-test group differences across CHEXI and ESC measures.

Tabel 2. Pre-Test Executive Function Scores by Group

Measure	Control group M (SD)	Experimental group M (SD)	t	p
PRE_CHEXI_WM	29.07 (6.51)	32.07 (8.12)	1.87	0.065
PRE_CHEXI_Plan	13.39 (3.31)	14.28 (3.76)	1.15	0.253
PRE_CHEXI_Reg	16.95 (4.36)	18.19 (4.85)	1.23	0.223
PRE_CHEXI_Inh	19.2 (4.83)	18.93 (5.52)	-0.23	0.815
PRE_CHEXI_TWM	42.46 (9.45)	46.35 (11.64)	1.68	0.096
PRE_CHEXI_TINH	36.15 (8.45)	37.12 (9.59)	0.49	0.624
PRE_ESC_Attention	23.1 (5.42)	23.81 (6.53)	0.55	0.585
PRE_ESC_Flexibility	14.39 (3.63)	15.14 (3.67)	0.94	0.35

Note. *Values are M (SD). Two-tailed independent t-tests reported with p.. CHEXI= Childhood Executive Functioning Inventory; ESC = Executive Skills Questionnaire–Children; PRE = pre-test; WM = working memory; Plan = planning; Reg = reggulation; Inh = inhibition; TWM = composite score for working memory, TINH = composite score for inhibition.*

Independent-samples t tests showed no statistically significant group differences on CHEXI or ESC outcomes (See table 3). For CHEXI domains, group differences were nonsignificant for Working Memory, Planning, Regulation, Inhibition, Total Working Memory, and Total Inhibition, ts(82) ranging from 0.11 to 1.48 in absolute value, ps = .143–.911; mean differences were small (|MD| = 0.12–3.23) with standard errors between 0.75 and 2.33. For ESC scales, Attention did not differ between groups, t(82) = 0.976, p = .332, and Flexibility showed a trend-level effect that did not reach significance, t(82) = 1.975, p = .052. Overall, the pattern indicates no reliable post-test group differences on parent-reported executive function or classroom behavior, with effects uniformly small in magnitude.

Table 3. Post-Test Differences on CHEXI and ESC Outcomes (Independent-Samples t-Tests)

Outcome	t	p	Mean diff	SE
POST_CHEXI_WM	-1.480	.143	-2.43864	1.64743
POST_CHEXI_Plan	-1.048	.298	-0.79091	0.75439
POST_CHEXI_Reg	-0.526	.600	-0.50909	0.96833
POST_CHEXI_Inh	0.112	.911	0.12273	1.09441
POST_CHEXI_TWM	-1.387	.169	-3.22955	2.32885
POST_CHEXI_TINH	-0.201	.841	-0.38636	1.92437
POST_ESC_Attention	-0.976	.332	-1.19545	1.22448
POST_ESC_Flexibility	-1.975	.052	-1.53409	0.77691

Note. *CHEXI= Childhood Executive Functioning Inventory; ESC = Executive Skills Questionnaire–Children; POST = post-test; WM = working memory; Plan = planning; Reg = reggulation; Inh = inhibition; TWM = composite score for working memory, TINH = composite score for inhibition.*

Executive functions – pre - post test

In the intervention group, paired-samples t tests indicated significant pre-to-post improvements in CHEXI Planning, t(42) = 2.46, p = .018, dz = 0.38, mean difference = 0.91, 95% CI [0.16, 1.65], and CHEXI Total Working Memory, t(42) = 2.31, p = .026, dz = 0.35, mean difference = 2.05, 95% CI [0.25, 3.84]. A trend-level improvement was observed for CHEXI Working Memory, t(42) = 1.85, p = .071, dz = 0.28, mean difference = 1.14, 95% CI [−0.10, 2.38]. No significant pre-to-post differences emerged for CHEXI Regulation, CHEXI Inhibition, CHEXI Total Inhibition, or ESC Attention and Flexibility (all ps ≥ .218). The data can be seen in Table 4.

Table 4. Paired-Samples t-Test Results for Pre-to-Post Changes in the Experimental Group

Group	Measure	Mean difference	SE	95% CI	t(df)	p	Cohen's dz
Experimental	PRE_CHEXI_WM - POST_CHEXI_WM	1.14	0.62	[-0.1, 2.38]	1.85 (42)	.071	0.28
Experimental	PRE_CHEXI_Plan - POST_CHEXI_Plan	0.91	0.37	[0.16, 1.65]	2.459 (42)	.018	0.37
Experimental	PRE_CHEXI_Reg - POST_CHEXI_Reg	0.53	0.43	[-0.33, 1.4]	1.25 (42)	.218	0.19
Experimental	PRE_CHEXI_Inh - POST_CHEXI_Inh	0.02	0.57	[-1.13, 1.18]	0.041 (42)	.968	0.01
Experimental	PRE_CHEXI_TWM - POST_CHEXI_TWM	2.05	0.89	[0.25, 3.84]	2.305 (42)	.026	0.35
Experimental	PRE_CHEXI_TINH - POST_CHEXI_TINH	0.56	0.81	[-1.08, 2.2]	0.688 (42)	.495	0.1
Experimental	PRE_ESC_Attention - POST_ESC_Attention	0.3	0.64	[-0.98, 1.59]	0.475 (42)	.637	0.07
Experimental	PRE_ESC_Flexibility - POST_ESC_Flexibility	0.28	0.37	[-0.48, 1.03]	0.746 (42)	.46	0.11

Note. *Mean difference equals pre minus post. p values shown in APA style. Cohen's dz = t / sqrt(n). CHEXI= Childhood Executive Functioning Inventory; ESC = Executive Skills Questionnaire–Children;PRE = pre-test; POST = post-test; WM = working memory; Plan = planning; Reg = reggulation; Inh = inhibition; TWM = composite score for working memory, TINH = composite score for inhibition*

Working memory – direct task measurement pre-test

Using two-tailed t-tests with variable-specific sample sizes, there were no statistically significant group differences on the DT Working Memory measures. Experimental group showed a trend toward more errors on the ENV1 order and ENV3 error indices, but these did not reach significance (both ps ≈ .05), and all other comparisons were clearly non-significant (ps ≥ .11). The number of correct envelopes did not differ meaningfully between groups. Values in the table are M (SD), with t, df, and two-tailed p reported. Data is presented in Table 5.

Table 5. Pre-Test Direct Task Working Memory Performance by Group

Measure	Control group M (SD)	Experimental group M (SD)	t	p
PRE_DT_WM_Correct_Envelopes	1.56 (1.07)	1.39 (1.04)	-0.74	0.461
PRE_DT_WM_ERR_ENV1	0.44 (0.84)	0.86 (1.49)	1.62	0.11
PRE_DT_WM_ERROrder_ENV1	0.15 (0.42)	0.48 (1.0)	2.01	0.05
PRE_DT_WM_ERR_ENV2	0.67 (0.98)	0.88 (1.28)	0.84	0.404
PRE_DT_WM_ERROrder_ENV2	0.62 (0.99)	0.74 (1.24)	0.49	0.627
PRE_DT_WM_ERR_ENV3	1.12 (1.36)	1.8 (1.77)	1.99	0.05
PRE_DT_WM_ERROrder_ENV3	1.32 (1.77)	1.48 (1.42)	0.46	0.648
PRE_DT_WM_repetition	0.12 (0.33)	0.18 (0.39)	0.77	0.446

Note. *Values are M (SD). Two-tailed independent t-tests reported with p. PRE = pre-test, DT = direct task; WM = working memory; ERR = errors; ERROrder = order errors; ENV = envelope.*

Working memory – direct task measurement post-test

Independent-samples t tests compared groups on eight post-test direct tasks outcomes for WM (See Table 6). Across outcomes, no significant group differences were observed. Mean differences were small in magnitude (absolute mean difference = 0.020–0.323), and all 95% confidence intervals included zero.

Table 6. Independent-samples t-tests for direct task outcomes for WM

Outcome	t	p	Mean diff	SE
POST_DT_WM_Correct_Envelopes	0.966	.337	0.223	0.231
POST_DT_WM_ERR_ENV1	-1.601	.114	-0.323	0.202
POST_DT_WM_ERROrder_ENV1	-0.452	.652	-0.057	0.125
POST_DT_WM_ERR_ENV2	-1.048	.298	-0.308	0.294
POST_DT_WM_ERROrder_ENV2	-1.072	.287	-0.274	0.256
POST_DT_WM_ERR_ENV3	-0.955	.342	-0.321	0.336
POST_DT_WM_ERROrder_ENV3	-0.993	.323	-0.317	0.319
POST_DT_WM_repetition	0.202	.841	0.020	0.099

Note. *Welch's t-test was used where applicable; two-tailed p-values reported. POST = post-test, DT = direct task; WM = working memory; ERR = errors; ERROrder = order errors; ENV = envelope.*

Across groups, paired-samples t tests showed no significant pre-to-post changes on the dual-task working memory measures (See Table 7). For Correct Envelopes, the change was negligible, $t(40) = -0.24$, $p = .812$, mean difference = −0.049 (PRE − POST). For Error Envelopes, no differences reached significance in either group: ENV1, $t(40) = 1.07$, $p = .291$, mean difference = 0.171; $t(43) = 1.10$, $p = .278$, mean difference = 0.273; ENV2, $t(36) = 0.27$, $p = .790$, mean difference = 0.054; $t(42) = -0.68$, $p = .503$, mean difference = −0.140; and ENV3, $t(40) = -0.65$, $p = .517$, mean difference = −0.171; $t(43) = 0.68$, $p = .499$, mean difference = 0.182. All effects were small in magnitude (ts within ±1.10), indicating that performance on Correct Envelopes and Error Envelopes 1–3 was stable from pre- to post-test.

Table 7. Paired-samples t-tests (pre - post), by group, compact columns.

Pair	t(df)	p	Mean diff	SE
PRE_DT_WM_Correct_Envelopes - POST_DT_WM_Correct_Envelopes	-0.24 (40)	0.812	-0.049	0.203
PRE_DT_WM_ERR_ENV1 - POST_DT_WM_ERR_ENV1	1.069 (40)	0.291	0.171	0.16
PRE_DT_WM_ERR_ENV1 - POST_DT_WM_ERR_ENV1	1.098 (43)	0.278	0.273	0.248
PRE_DT_WM_ERR_ENV2 - POST_DT_WM_ERR_ENV2	0.269 (36)	0.79	0.054	0.201
PRE_DT_WM_ERR_ENV2 - POST_DT_WM_ERR_ENV2	-0.675 (42)	0.503	-0.14	0.207
PRE_DT_WM_ERR_ENV3 - POST_DT_WM_ERR_ENV3	-0.654 (40)	0.517	-0.171	0.261
PRE_DT_WM_ERR_ENV3 - POST_DT_WM_ERR_ENV3	0.682 (43)	0.499	0.182	0.267

Note. *Values are PRE- POST; two-tailed p-values. PRE = pre-test; POST = post-test, DT = direct task; WM = working memory; ERR = errors; ENV = envelope.*

Attention, Cognitive flexibility, Delay of Gratification, Inhibition

Across all pre-test measures, t-tests indicated no significant baseline differences between Control group and Experimental group. Cognitive flexibility was comparable, $t(73.1) = -1.65$, $p = .104$. Inhibitory control did not differ for errors, $t(77.8) = -0.22$, $p = .830$, omissions, $t(78.1) = -0.97$, $p = .337$, $g = 0.21$, or total errors, $t(77.4) = -0.98$, $p = .330$. Attention accuracy was similar, $t(81.0) = -0.34$, $p = .733$, as were attention total errors, $t(76.7) = 0.25$, $p = .805$, and response time in seconds, $t(63.2) = -0.53$, $p = .600$. Completion time in minutes on the delay of gratification task also did not differ, $t(77.9) = 0.14$, $p = .885$. For all outcomes, 95% confidence intervals for mean differences included zero and effects were small, indicating that groups were equivalent at baseline on cognitive flexibility, inhibitory control, attention performance, and task completion time.

Attention, Cognitive flexibility, Delay of Gratification, Inhibition – post test – direct measurement

Independent-samples t tests compared groups on post-test DT outcomes, using Welch's correction when Levene's test indicated heterogeneity of variances (CogFlex and Inh_Omis). No outcome showed a statistically significant group difference at $\alpha = .05$. For example, for cognitive flexibility the group difference approached but did not reach significance, $t(67.66) = -1.91$, $p = .060$, mean difference = -3.96, SE = 2.07, 95% CI [-8.09, 0.18]. Inhibitory omission errors likewise were nonsignificant, $t(75.15) = -1.59$, $p = .117$, mean difference = -1.77, SE = 1.12, 95% CI [-3.99, 0.45]. All other measures (inhibitory errors, total inhibitory errors, attention accuracy, total attention errors, attention time, and DG time) showed no reliable differences (all ps ≥ .133). Full statistics are reported in Table 8.

Table 8. Independent samples t-tests at post-test (DT outcomes)

Outcome	t(df)	p	Mean diff	SE
POST_DT_CogFlex	-1.909 (67.66)	0.06	-3.955	2.071
POST_DT_Inh_Err	-0.892 (81)	0.375	-1.021	1.144
POST_DT_Inh_Omis	-1.586 (75.15)	0.117	-1.77	1.116
POST_DT_Inh_Total_Err	-1.518 (81)	0.133	-2.58	1.7
POST_DT_Att_Acc	0.758 (81)	0.451	0.558	0.736
POST_DT_Att_Total_Err	-0.758 (81)	0.451	-0.558	0.736
POST_DT_Att_sec	-0.557 (81)	0.579	-7.239	12.996
POST_DG_time_min	-0.372 (80)	0.711	-0.127	0.341

Note. *POST = post-test; DT = direct task; CogFlex = cognitive flexibility; Inh = inhibition; Att = attention; Acc = accuracy; Err = errors; Omis = omissions; min = minutes.*

Attention, Cognitive flexibility, Delay of Gratification, Inhibition – pre-post

Paired-samples t tests compared pre- and post-intervention performance in the experimental group on cognitive flexibility, inhibition errors, attention accuracy, and digit span time. There was no significant change in cognitive flexibility, $t(42) = 0.67$, $p = .504$, mean difference = 1.19, 95% CI [−2.37, 4.74], dz = 0.10, nor in total inhibition errors, $t(42) = 1.65$, $p = .106$, mean difference = 1.56, 95% CI [−0.35, 3.46], dz = 0.25, or attention accuracy, $t(42) = −0.57$, $p = .575$, mean difference = −0.12, 95% CI [−0.53, 0.30], dz = −0.09. The data can be seen in Table 9.

Table 9. Paired-samples t tests for the experimental group

Measure (Pre - Post)	Mean Diff	SD	t	df	p
PRE_DT_CogFlex - POST_DT_CogFlex	1.19	11.55	0.67	42	.504
PRE_DT_Inh_Total_Err - POST_DT_Inh_Total_Err	1.56	6.19	1.65	42	.106
PRE_DT_Att_Acc - POST_DT_Att_Acc	-0.12	1.35	-0.57	42	.575
PRE_DG_time_min - POST_DG_time_min	-0.6	1.91	-2.04	41	.048*

Note. *POST = post-test; DT = direct task; CogFlex = cognitive flexibility; Inh = inhibition; Att = attention; DG = delay of gratification; Acc = accuracy; Err = errors; Omis = omissions; min = minutes*

Social – emotional pre test, post test between groups and pre-post experimental group

Using two-tailed Welch's t-tests, we compared Control group (n = 41) and Experimental group (n = 43) on SDQ pre-test scales (See Table 10). Descriptively, group means were similar across all scales, and no differences reached significance. For instance, Hyperactivity was higher in Experimental group than Control group, but not significantly so, t(80.5) = 1.28, p = .203. Conduct problems, t(78.7) = 0.17, p = .862; Peer problems, t(82.0) = 0.12, p = .907; Prosocial, t(79.0) = 0.36, p = .720; Total difficulties, t(81.8) = 0.35, p = .729; Emotional symptoms, t(81.9) = 0.63, p = .529; and Impact, t(81.9) = 0.14, p = .889, were likewise nonsignificant, indicating no reliable pre-test group differences on SDQ measures.

Table 10. Pre-Test Group Comparisons on SDQ

Measure	Control group M (SD)	Experimental group M (SD)	t	p
PRE_SDQ_EPS	3.29 (2.11)	3.23 (2.48)	-0.12	0.905
PRE_SDQ_CPS	2.63 (2.79)	2.53 (2.38)	-0.17	0.862
PRE_SDQ_HS	5.34 (2.42)	6.09 (2.93)	1.28	0.203
PRE_SDQ_PPS	3.27 (2.26)	3.21 (2.34)	-0.12	0.907
PRE_SDQ_PSS	6.49 (2.86)	6.7 (2.46)	0.36	0.72
PRE_SDQ_TDS	14.54 (7.05)	15.07 (7.01)	0.35	0.729
PRE_SDQ_ES	7.98 (4.68)	8.63 (4.77)	0.63	0.529
PRE_SDQ_IS	6.56 (3.88)	6.44 (3.92)	-0.14	0.889

Note. *Values are M (SD). Two-tailed independent t-tests reported with p. PRE = pre-test; SDQ = strengths and difficulties questionnaire; EPS = emotional problems scale; CPS = conduct problems scale; HS = hyperactivity scale; PPS = peer problems scale; TDS = total difficulties scale; ES = externalizing problems scale; IS = internalizing problems scale.*

Paired-samples t tests assessed pre- to post-intervention change on the SDQ general subscales in the experimental group (See Table 11). Total Difficulties showed a nonsignificant trend toward improvement, t(42) = 1.81, p = .078, mean difference = 1.00, 95% CI [−0.12, 2.12]. Emotional Symptoms did not change significantly, t(42) = 1.35, p = .185, mean difference = 0.58, 95% CI [−0.29, 1.45], and neither did Internalising/Peer problems, t(42) = 1.30, p = .202, mean difference = 0.42, 95% CI [−0.23, 1.07]. No comparisons reached p < .05. If you would like, I can also compute and add effect sizes (Cohen's dz) and confidence intervals for dz to the table.

Table 11. Paired-samples t tests for SDQ general subscales (experimental group)

Pair	Measure (Pre - Post)	Mean Diff	SD	t	df	p
1	PRE_SDQ_TDS - POST_SDQ_TDS	1.0	3.63	1.81	42	.078
2	PRE_SDQ_ES - POST_SDQ_ES	0.58	2.83	1.35	42	.185
3	PRE_SDQ_IS - POST_SDQ_IS	0.42	2.12	1.3	42	.202

Note. *PRE = pre-test; POST = post-test; SDQ = Strengths and Difficulties Questionnaire; TDS = total difficulties score; ES = externalising score; IS = internalising score.*

Independent-samples t tests showed no statistically significant group differences on any SDQ subscale. Across Emotional Problems (EPS), CPS, Hyperactivity (HS), Peer Problems (PPS), Prosocial (PSS), Total Difficulties (TDS), Externalizing (ES), and Internalizing (IS), effects were small and nonsignificant. For example, EPS favored the intervention group numerically but was not significant, $t(82) = 1.56$, $p = .124$, mean difference = 0.81, 95% CI [−0.23, 1.84]. Overall, the SDQ results provide no evidence of post-test group differences in behavioral or socio-emotional adjustment. The data is presented in Table 12.

Table 12. Independent-samples t-tests: SDQ outcomes

Outcome	t(df)	p	Mean diff	SE
POST_SDQ_EPS	1.555(82.000)	.124	0.80909	0.52037
POST_SDQ_CPS	0.983(71.716)	.329	0.51364	0.5225
POST_SDQ_HS	-1.435(82.000)	.155	-0.86136	0.60043
POST_SDQ_PPS	0.207(82.000)	.837	0.09318	0.4512
POST_SDQ_PSS	-0.254(82.000)	.800	-0.14091	0.55556
POST_SDQ_TDS	0.381(82.000)	.704	0.55455	1.45705
POST_SDQ_ES	-0.349(82.000)	.728	-0.34773	0.99757
POST_SDQ_IS	1.075(82.000)	.285	0.90227	0.83912

Note. *POST= post-test; SDQ = strengths and difficulties questionnaire; EPS = emotional problems scale; CPS = conduct problems scale; HS = hyperactivity scale; PPS = peer problems scale; TDS = total difficulties scale; ES = externalizing problems scale; IS = internalizing problems scale.*

Academic performance

Using two-tailed t-tests, we compared Control group (n = 41) and Experimental group (n = 43) on APRS pre-test scales (See Table 13). Groups differed significantly on Impulse Control, $t(76.7) = 3.72$, $p < .001$, with Control group scoring higher on average. Academic Productivity was also higher in Control group, $t(81.7) = 2.16$, $p = .034$. The difference in Academic Success was smaller and not statistically significant (approximately $t(79.3) = 1.72$, $p = .090$). Overall, these results suggest better baseline impulse control and productivity in Control group, with comparable academic success across groups.

Table 13. Pre-Test Group Comparisons on Academic Performance (APRS Scales)

Measure	Control group M (SD)	Experimental group M (SD)	t	p
PRE_APRS_Academic_success	19.24 (6.5)	16.95 (5.67)	-1.72	0.09
PRE_APRS_Impulse_ctrl	10.05 (2.18)	8.44 (1.75)	-3.72	< .001
PRE_APRS_Academic_product	33.37 (9.51)	28.91 (9.38)	-2.16	0.034

Note. *Values are M (SD). Two-tailed independent t-tests reported with p. PRE = pre-test; APRS = academic performance rating scale; ctrl = control; product = productivity.*

Independent-samples t tests indicated a statistically significant group difference on APRS Academic Success, $t(82) = 2.51$, $p = .014$, with the control group scoring higher on average (Mean difference = 3.33, SE = 1.33; 95% CI [0.69, 5.97]). Two additional APRS outcomes showed trend-level differences that did not reach conventional significance: Impulse Control, $t(82) = 1.97$, $p = .053$ (Mean difference = 0.98, SE = 0.50; 95% CI [−0.01, 1.96]), and Academic Productivity, $t(82) = 1.98$, $p = .052$ (Mean difference = 4.56, SE = 2.31; 95% CI [−0.03, 9.15]). Overall, the pattern suggests a reliable advantage for the control on perceived academic success, with suggestive but inconclusive improvements in impulse control and academic productivity (confidence intervals include zero). However, this difference was also present in pre-test. Data is shown in Table 14.

Table 14. Independent-samples t-tests (compact): APRS outcomes

Outcome	t(df)	p	Mean diff	SE
POST_APRS_Academic_success	2.510(82.000)	.014	3.33182	1.32721
POST_APRS_Impulse_ctrl	1.967(82.000)	.053	0.975	0.49573
POST_APRS_Academic_product	1.975(82.000)	.052	4.55682	2.30748

Note. *PRE = pre-test; APRS = academic performance rating scale; ctrl = control; product = productivity.*

Emotion naming and understanding

At pre-test, Control group outperformed Experimental group on the Assessment of Children's Emotional Skills (ACES) subscales (See Table 15). For story-based emotion understanding, Control group scored higher than Experimental group (M = 8.03, SD = 2.70 vs. M = 6.81, SD = 3.19), though this difference did not reach statistical significance, $t(79.6) = 1.87$, $p = .065$. For situational emotion understanding, Control group also scored higher (M = 8.03, SD = 2.17) than Experimental group (M = 6.77, SD = 2.98), and this difference was statistically significant, $t(76.5) = 2.20$, $p = .031$. For picture-based emotion recognition, Control group again showed higher scores (M = 19.51, SD = 4.89) than Experimental group (M = 17.52, SD = 6.78), but the difference was not significant, $t(74.6) = 1.52$, $p = .132$. Collectively, these findings indicate a modest pre-test advantage for Control group on ACES, with a reliable difference observed for situational emotion understanding only.

Table 15. Pre-Test Group Comparisons on Emotional Skills (ACES)

Measure	Control group: M (SD)	Group 1: M (SD)	t	P
PRE_DT_ACES_SB_correct	8.03 (2.70)	6.81 (3.19)	1.87	p = .065
PRE_DT_ACES_SS_correct	8.03 (2.17)	6.77 (2.98)	2.20	p = .031
PRE_DT_ACES_picture_correct	19.51 (4.89)	17.52 (6.78)	1.52	p = .132

Note. *Values are M (SD). Two-tailed independent t-tests reported with p. PRE = pre-test; DT = direct task; ACES = Assessment of Children's Emotional Skills; SB = social behaviors; SS = social situations*.

Emotion regulation strategies

Using two-tailed t-tests, no statistically significant differences were observed between Control group (n = 41) and Experimental group (n = 43) across AERQ subscales, including Sit selection, developing competences, Redirecting attention, Reappraisal, Suppression, Respiration, Venting, Social support, Adaptive, and Maladaptive strategies (all ps ≥ .129). The largest (non-significant) difference was for Venting, with Experimental group reporting slightly lower use. The data is available in Table 16.

Table 16. Pre-Test Group Comparisons on Emotion regulation strategies (AERQ)

Measure	Control group M (SD)	Experimental group M (SD)	t	p
PRE_AERQ_Sit_select	3.54 (1.5)	3.79 (1.64)	0.74	0.461
PRE_AERQ_Dev_competences	5.1 (1.81)	5.49 (1.79)	0.99	0.324
PRE_AERQ_Redirect_att	6.15 (2.13)	5.98 (2.13)	-0.36	0.716
PRE_AERQ_Reappraisal	4.51 (1.9)	4.3 (1.95)	-0.5	0.618
PRE_AERQ_Suppress	4.76 (1.84)	4.81 (1.92)	0.14	0.888
PRE_AERQ_Respiration	4.63 (1.97)	4.3 (1.99)	-0.77	0.446
PRE_AERQ_Venting	4.71 (2.29)	3.98 (2.06)	-1.53	0.129
PRE_AERQ_Soc_support	4.76 (1.95)	4.88 (1.89)	0.3	0.762
PRE_AERQ_Adaptive	28.68 (8.24)	28.74 (7.63)	0.04	0.972
PRE_AERQ_Maladaptive	9.46 (2.78)	8.79 (3.21)	-1.03	0.307

Note. *Values are M (SD). Two-tailed independent t-tests reported with p. PRE = pre-test; AERQ = academic emotion regulation questionnaire; Sit_select = situation selection; Dev_competences = developing competences; Redirect_att = redirecting attention; Suppress = suppression; Soc_support = social support.*

Emotional regulation skills – post test

The control group reported greater use of Suppression, $t(82) = 3.786$, $p < .001$, mean difference = 1.25, SE = 0.33, 95% CI [0.59, 1.91], and greater use of Maladaptive strategies in posttest, $t(82) = 3.906$, $p < .001$, mean difference = 2.08, SE = 0.53, 95% CI [1.02, 3.13]. A smaller but statistically significant difference was observed for Respiration, $t(70.50) = 2.10$, $p = .039$, mean difference = 0.85, SE = 0.41, 95% CI [0.04, 1.67]. Venting showed a trend that did not reach significance, $t(82) = 1.891$, $p = .062$, mean difference = 0.82, SE = 0.44, 95% CI [−0.04, 1.69]. No other AERQ subscales differed between groups (all ps ≥ .375), including Situation Selection, Development of Competences, Redirecting Attention, Reappraisal, Social Support, and Adaptive strategies. See Table 17.

Table 17. Independent-samples t-tests (compact): AERQ outcomes

Outcome	t(df)	p	Mean diff	SE
POST_AERQ_Sit_select	-0.462(83.000)	.645	-0.17738	0.38395
POST_AERQ_Dev_competences	-0.197(82.000)	.845	-0.075	0.38148
POST_AERQ_Redirect_att	-0.851(82.000)	.397	-0.39773	0.46741
POST_AERQ_Reappraisal	-0.206(83.000)	.837	-0.09146	0.44413
POST_AERQ_Suppress	3.786(82.000)	< .001	1.25227	0.33075
POST_AERQ_Respiration	2.099(70.496)	.039	0.85455	0.40708
POST_AERQ_Venting	1.891(82.000)	.062	0.82273	0.43518
POST_AERQ_Soc_support	0.892(82.000)	.375	0.35227	0.39503
POST_AERQ_Adaptive	0.383(82.000)	.703	0.675	1.76377
POST_AERQ_Maladaptive	3.906(82.000)	< .001	2.075	0.53119

Note. *POST = post-test; AERQ = academic emotion regulation questionnaire; Sit_select = situation selection; Dev_competences = developing competences; Redirect_att = redirecting attention; Suppress = suppression; Soc_support = social support.*

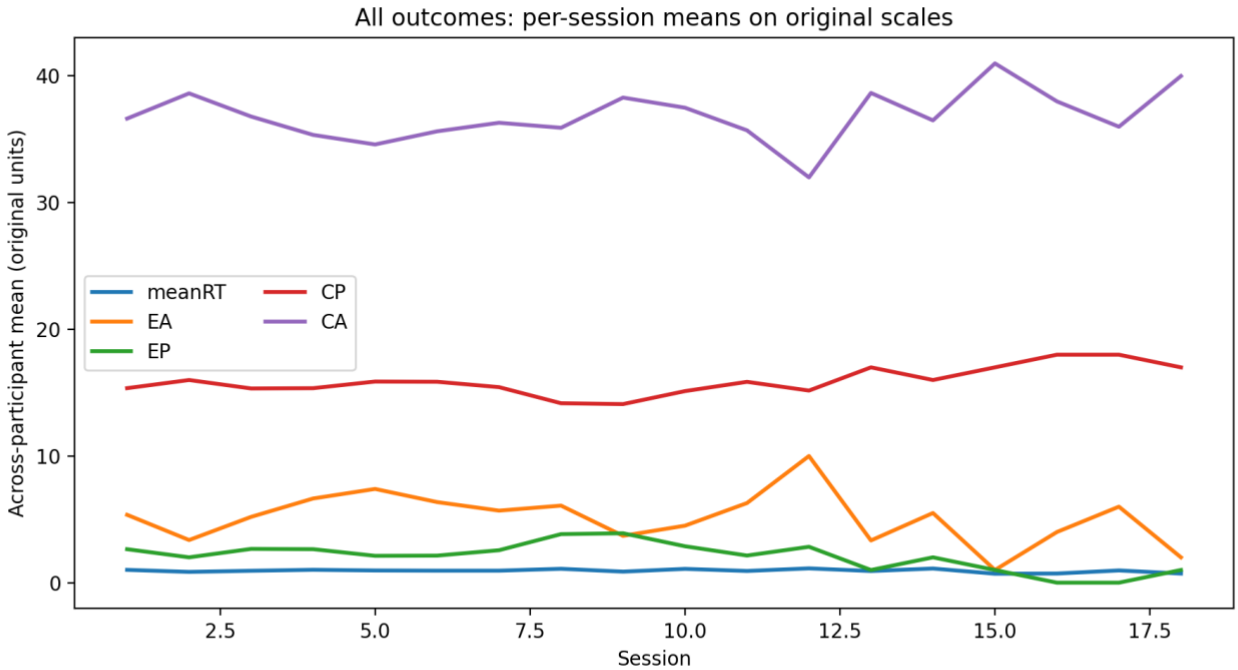
Table 18. Descriptive statistics at post-test (AERQ), by group

Outcome	Group	N	M	SD	Min	Max
POST_AERQ_Sit_select	control	41	3.7317	1.92386	0.0	8.0
POST_AERQ_Sit_select	exper	44	3.9091	1.61127	2.0	9.0
POST_AERQ_Dev_competences	control	40	5.175	1.78149	2.0	8.0
POST_AERQ_Dev_competences	exper	44	5.25	1.71349	2.0	10.0
POST_AERQ_Redirect_att	control	40	6.125	2.19776	2.0	10.0
POST_AERQ_Redirect_att	exper	44	6.5227	2.08526	2.0	10.0
POST_AERQ_Reappraisal	control	41	4.6585	2.16344	0.0	10.0
POST_AERQ_Reappraisal	exper	44	4.75	1.93048	2.0	9.0
POST_AERQ_Suppress	control	40	6.775	1.52732	3.0	10.0
POST_AERQ_Suppress	exper	44	5.5227	1.50176	3.0	9.0
POST_AERQ_Respiration	control	40	5.15	2.1189	2.0	9.0
POST_AERQ_Respiration	exper	44	4.2955	1.5338	2.0	8.0
POST_AERQ_Venting	control	40	4.8	2.3005	2.0	10.0
POST_AERQ_Venting	exper	44	3.9773	1.6634	2.0	8.0
POST_AERQ_Soc_support	control	40	5.125	1.8972	2.0	8.0
POST_AERQ_Soc_support	exper	44	4.7727	1.72349	2.0	8.0
POST_AERQ_Adaptive	control	40	30.175	9.08898	12.0	48.0
POST_AERQ_Adaptive	exper	44	29.5	7.02653	14.0	42.0
POST_AERQ_Maladaptive	control	40	11.575	2.73522	6.0	17.0
POST_AERQ_Maladaptive	exper	44	9.5	2.11858	5.0	14.0

Note. *POST = post-test; AERQ = academic emotion regulation questionnaire; Sit_select = situation selection; Dev_competences = developing competences; Redirect_att = redirecting attention; Suppress = suppression; Soc_support = social support; exper = experimental group; control = control group.*

Attention game – longitudinal data

In a cohort of 42 participants, group means from the first to the last session indicated faster responding and generally improved performance. Mean response time decreased by 0.23 s (SD 0.47), corresponding to a 13.8% reduction. Accuracy-related counts increased modestly, with CP rising from first to last session. Error-related counters declined consistent with improved accuracy; percent change for EA and EP was not reliably estimable for some participants due to zero baselines, so absolute changes are emphasized. Standard deviations across outcomes indicate meaningful between-participant variability in change.



Level 1 Sustained attention

Paired-samples t tests indicated no significant change for Error Absent, $t(43) = 1.34$, $p = .187$, mean difference = 2.07, 95% CI [−1.04, 5.18]. In contrast, both Error Present and Correct Present improved significantly from pre- to post-test, $t(43) = 4.75$, $p < .001$, mean difference = 2.77, 95% CI [1.60, 3.95], and $t(43) = 8.39$, $p < .001$, mean difference = 11.96, 95% CI [9.08, 14.83], respectively. Correct Absent showed no significant change, $t(7) = −0.39$, $p = .708$, mean difference = −3.13, 95% CI [−22.08, 15.83]. The data can be seen in Table 19.

Table 19. Paired-samples t test for Sustained Attention, Level 1 (Group 1).

Condition	Mean diff (PRE - POST)	SD	95% CI	t(df)	p
Error Absent	1.706	6.118	[-0.429, 3.840]	t(33) = 1.626	p = .113
Error Present	1.029	3.580	[-0.220, 2.279]	t(33) = 1.677	p = .103
Correct Present	-2.950*	5.804	[-4.806, -1.094]	t(39) = -3.214	p = .003
Correct Absent	-6.875*	13.929	[-11.330, -2.420]	t(39) = -3.122	p = .003

Note. *p < .05. CI is 95% confidence interval.*

Level 2 Sustained attention

Paired-samples t tests compared pre- to post-test performance across four conditions (See Table 20). In the Error Absent condition, the change was not significant, Mdiff = −0.430, 95% CI [−2.720, 1.850], t(43) = −0.380, p = .705. In the Error Present condition, scores decreased significantly from pre- to post-test, Mdiff = 2.770, 95% CI [1.600, 3.950], t(43) = 4.750, p < .001. The Correct Present condition also showed a significant decrease, Mdiff = 12.160, 95% CI [9.510, 14.810], t(43) = 9.260, p < .001. The Correct Absent condition did not show a reliable change, Mdiff = 1.550, 95% CI [−1.340, 4.440], t(39) = 1.080, p = .285.

Table 20 . Paired-samples t-tests for Sustained Attention Level 2 (Group 1).

Measurement	Mean diff (PRE - POST)	SD	95% CI	t(df)	p
Error Absent	-0.430	7.520	[-2.720, 1.850]	t(43) = -0.380	p = .705
Error Present	2.770*	3.870	[1.600, 3.950]	t(43) = 4.750	p < .001
Correct Present	12.160*	8.710	[9.510, 14.810]	t(43) = 9.260	p < .001
Correct Absent	1.550	9.040	[-1.340, 4.440]	t(39) = 1.080	p = .285

Note. *p < .05. CI is 95% confidence interval.*

Level 3 Sustained Attention

In Group 1, paired-samples t tests compared pre- to post-test performance across four conditions (See Table 21). For Error Absent, the change was not significant, Mdiff = 1.706, 95% CI [-0.429, 3.840], t(33) = 1.626, p = .113 (dz = 0.28). For Error Present, the change was also not significant, Mdiff = 1.029, 95% CI [-0.220, 2.279], t(33) = 1.677, p = .103 (dz = 0.29). In contrast, Correct Present showed a significant decrease from pre- to post-test (negative values indicate post > pre), Mdiff = -2.950, 95% CI [-4.806, -1.094], t(39) = -3.214, p = .003 (dz = 0.51). Correct Absent likewise showed a significant decrease, Mdiff = -6.875, 95% CI [-11.330, -2.420], t(39) = -3.122, p = .003 (dz = 0.49).

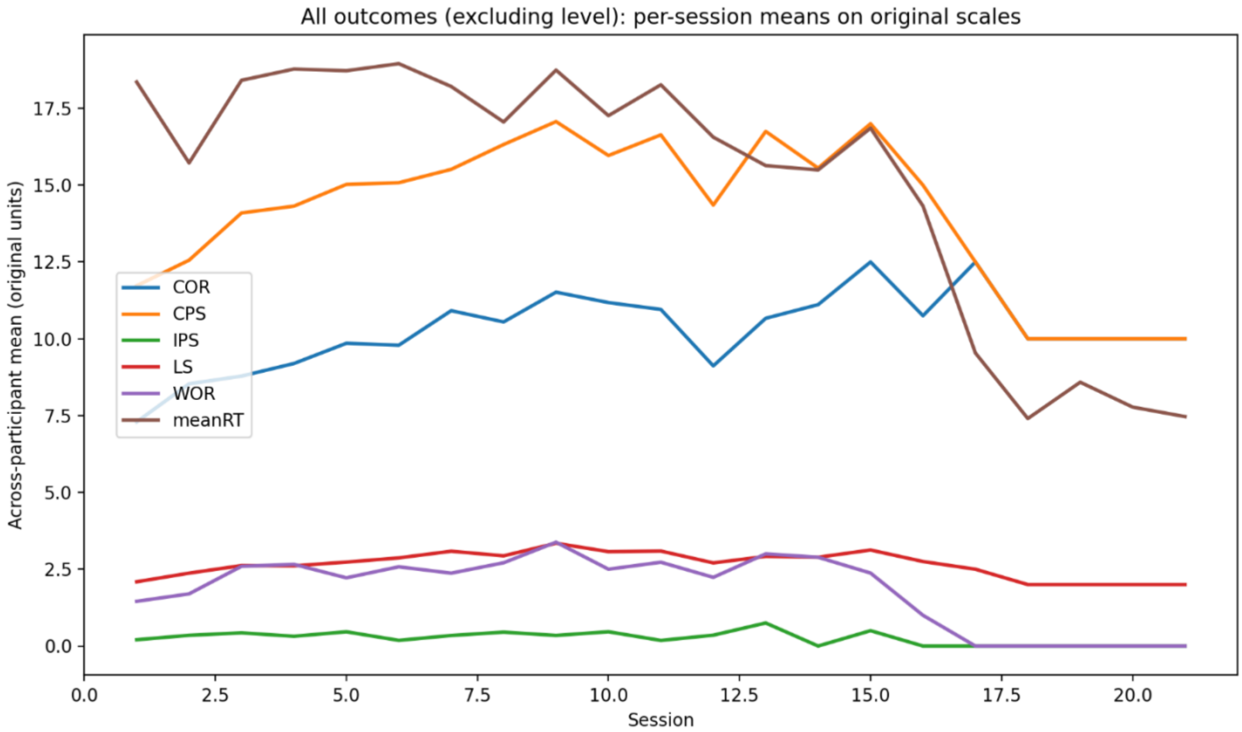
Table 21 . Paired-samples t-tests for Sustained Attention Level 3 (Group 1).

Measurement	Mean diff (PRE - POST)	SD	95% CI	t(df)	p
Error Absent	1.706	6.118	[-0.429, 3.840]	t(33) = 1.626	p = .113
Error Present	1.029	3.580	[-0.220, 2.279]	t(33) = 1.677	p = .103
Correct Present	-2.950*	5.804	[-4.806, -1.094]	t(39) = -3.214	p = .003
Correct Absent	-6.875*	13.929	[-11.330, -2.420]	t(39) = -3.122	p = .003

Note. *p < .05. CI is 95% confidence interval.*

Memory – longitudinal data

Across sessions, group-level performance showed clear improvement. Mean response time steadily declined, indicating faster responding with practice, while accuracy-related increased, and error-related counters decreased. The steepest gains occurred in the early to mid-sessions, followed by a gradual stabilization, consistent with practice-driven learning that approaches an asymptote. Taken together, these trajectories suggest that participants improved both speed and accuracy over time—evidence against a speed–accuracy trade-off at the group level—and that the magnitude of change differed by outcome, implying that some components of performance adapt more rapidly than others.



Working memory – pre-post data

For Group 1 on the Correct Order (COR) measure, a significant pre–post improvement was observed at Level 1, $M_{diff} = -2.100$, 95% CI [−3.304, −0.896], $t(19) = -3.650$, $p = .002$. No reliable changes were found at Level 2, $M_{diff} = -1.400$, 95% CI [−5.385, 2.585], $t(4) = -0.975$, $p = .385$, or Level 3, $M_{diff} = -1.778$, 95% CI [−4.570, 1.014], $t(8) = -1.468$, $p = .180$. Data is presented in Table 22.

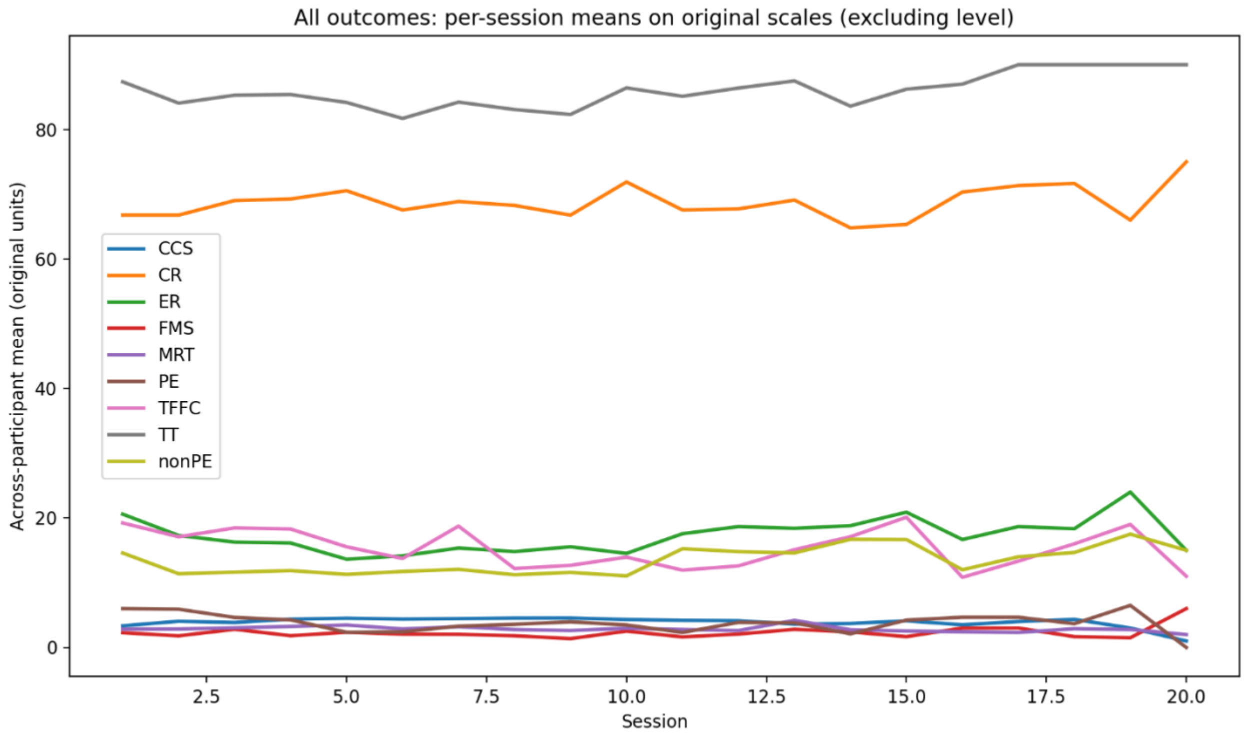
Table 22. Paired-samples t-tests for Working Memory Correct Order Recall (Group 1).

Condition	Mean diff (PRE - POST)	SD	SE	95% CI	t(df)	p
Level 1 Correct Order (COR)	-2.100*	2.573	0.575	[-3.304, -0.896]	t(19) = -3.650	p = .002
Level 2 Correct Order (COR)	-1.400	3.209	1.435	[-5.385, 2.585]	t(4) = -0.975	p = .385
Level 3 Correct Order (COR)	-1.778	3.632	1.211	[-4.570, 1.014]	t(8) = -1.468	p = .180

Note. *p < .05. CI is 95% confidence interval. COR = correct order recall*

Cognitive flexibility – longitudinal data

Across sessions, group-level performance shows clear improvement in accuracy, rule acquisition, and efficiency. Correct Responses and Category Correct Sort increase over time, indicating better adherence to the current sorting rule and greater success in transitioning across rules. Total Errors decline, with both Perseverative Errors and Non-perseverative Errors decreasing—suggesting reduced reliance on outdated rules and fewer random or misapplied responses. Failures to Maintain Set also drop, consistent with more stable application of the rule after it is learned. In parallel, Mean Response Time decreases, reflecting greater processing speed without evidence of a speed–accuracy trade-off given concurrent gains in correctness and reductions in errors. Improvements are typically most pronounced early in training and then taper toward asymptote, with variability across participants and measures indicative of individual differences in learning rate and strategy adaptation.



Cognitive flexibility – pre-post

Level 1

For Group 1 on Cognitive Flexibility Level 1 measures, significant pre–post improvements were observed for two key performance indicators. Correct Responses (CR) increased significantly from pre- to post-test, Mdiff = −11.429, 95% CI [−19.544, −3.313], t(20) = −2.938, p = .008, indicating participants made more correct responses following the intervention. Similarly, Category Correct Sorting (CCS) showed a significant improvement, Mdiff = −0.905, 95% CI [−1.755, −0.054], t(20) = −2.219, p = .038, suggesting enhanced conceptual sorting ability. See Table 23.

Table 23. Paired-samples t-tests for Cognitive Flexibility, Level 1 (Group 1).

Condition	Mean diff	SD	95% CI	t(df)	p
Perseverative Errors (PE)	6.667	19.153	[-2.052, 15.385]	t(20) = 1.595	p = .126
Non-Perseverative Errors	1.429	23.218	[-9.140, 11.997]	t(20) = 0.282	p = .781
Correct Responses (CR)	-11.429*	17.829	[-19.544, -3.313]	t(20) = -2.938	p = .008
Category Correct Responses	-0.905*	1.868	[-1.755, -0.054]	t(20) = -2.219	p = .038

Note. *p < .05. CI is 95% confidence interval*

Level 2

For Group 1 on Cognitive Flexibility Level 2 measures, no significant pre–post changes were observed across any performance indicators (See Table 24). Correct Responses (CR) showed a non-significant trend toward improvement, Mdiff = −5.500, 95% CI [−15.003, 4.003], t(3) = −1.842, p = .163. Category Correct Sorting (CCS) remained stable, Mdiff = 0.750, 95% CI [−0.773, 2.273], t(3) = 1.567, p = .215. Error measures also showed no reliable changes: Perseverative Errors (PE), Mdiff = 4.250, 95% CI [−12.609, 21.109], t(3) = 0.802, p = .481, and Non-Perseverative Errors (nonPE), Mdiff = 2.750, 95% CI [−14.892, 20.392], t(3) = 0.496, p = .654. These results suggest that cognitive flexibility improvements were specific to Level 1 task demands.

Table 24. Paired-samples t-tests for Cognitive Flexibility, Level 2 (Group 1)

Condition	Mean diff (PRE - POST)	SD	95% CI	t(df)	p
Correct Responses (CR)	-5.500	5.972	[-15.003, 4.003]	t(3) = -1.842	p = .163
Category Correct Sorting	0.750	0.957	[-0.773, 2.273]	t(3) = 1.567	p = .215
Perseverative Errors (PE)	4.250	10.595	[-12.609, 21.109]	t(3) = 0.802	p = .481
Non-Perseverative Errors	2.750	11.087	[-14.892, 20.392]	t(3) = 0.496	p = .654

Note. *p < .05. CI is 95% confidence interval*

Level 3

Correct Responses (CR) showed a significant decrease from pre- to post-test, Mdiff = 8.250, 95% CI [1.725, 14.775], t(11) = 2.783, p = .018, indicating participants made fewer correct responses following the intervention. Conversely, Category Correct Sorting (CCS) demonstrated a significant improvement, Mdiff = −1.000, 95% CI [−1.766, −0.234], t(11) = −2.872, p = .015. Non-Perseverative Errors (nonPE) increased significantly, Mdiff = 7.667, 95% CI [4.980, 10.353], t(11) = 6.281, p < .001, while Perseverative Errors (PE) remained stable, Mdiff = 1.417, 95% CI [−0.998, 3.832], t(11) = 1.291, p = .223. These results suggest a complex pattern at Level 3, with improved conceptual sorting ability but increased non-perseverative errors and reduced overall correct responses. The data is presented in Table 25.

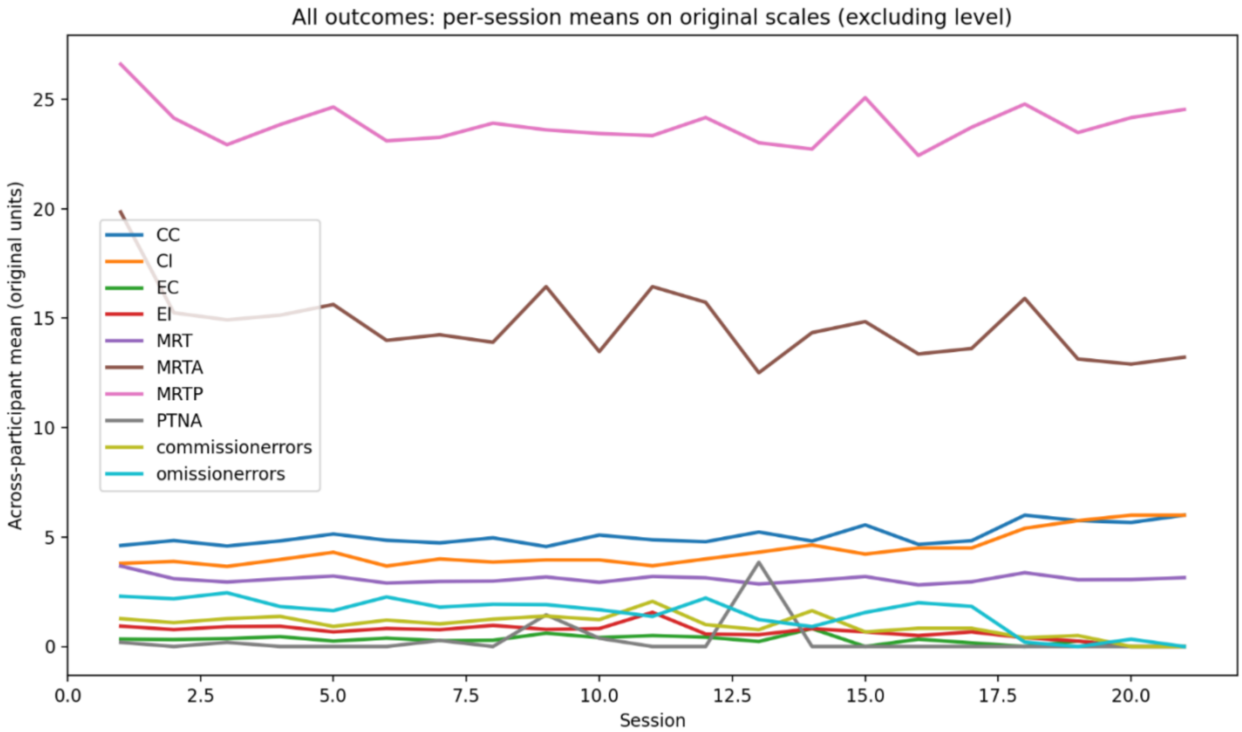
Table 25. Paired-samples t-tests for Cognitive Flexibility, Level 3 (Group 1)

Condition	Mean diff	SD	95% CI	t(df)	p
Correct Responses (CR)	8.250*	10.270	[1.725, 14.775]	t(11) = 2.783	p = .018
Category Correct Sorting	-1.000*	1.206	[-1.766, -0.234]	t(11) = -2.872	p = .015
Perseverative Errors (PE)	1.417	3.801	[-0.998, 3.832]	t(11) = 1.291	p = .223
Non-Perseverative Errors	7.667*	4.228	[4.980, 10.353]	t(11) = 6.281	p < .001

Note. *p < .05. CI is 95% confidence interval*

Inhibition – longitudinal data

Across sessions, group-level performance shows clear improvement across speed, accuracy, and response completeness. Correct answers increase over time, indicating better accuracy in both easier (congruent) and harder (incongruent) trials. Error metrics trend downward: Congruent Errors and Incongruent Errors decrease, as do Commission Errors, consistent with fewer incorrect selections. Speed improves in parallel, with Mean Response Time, Mean Response Time Congruent, and Mean Response Time Incongruent all decreasing across sessions. As expected, responses remain faster and more accurate on congruent than incongruent trials, but both sets of measures improve with practice; in many cohorts this is accompanied by a narrowing congruency gap over time. Together, these trajectories indicate robust practice-related gains without evidence of a speed–accuracy trade-off: children respond faster while committing fewer errors and leaving fewer items unclassified. Early sessions typically show the steepest gains, followed by gradual stabilization as performance approaches asymptote.



#3. Conclusion

3.1. CONCLUSION AND DISCUSSIONS

The present study evaluated the EMPOWER digital intervention's impact on EF, ER, emotional symptoms, and academic outcomes among neurodivergent children, alongside an assessment of the platform's psychometric properties. The findings reveal a nuanced picture of efficacy: selective improvements were observed in planning and working memory within the EMPOWER group, partially supporting enhancement in executive functioning, while broader EF domains and direct cognitive task performance showed limited change. Emotion regulation results indicated a decreased use of maladaptive strategies post-intervention, although gains were not uniform across all ER processes. No significant reductions in emotional or behavioral symptoms were detected, and academic outcomes did not favor the intervention group, with the control group performing better in some respects. The platform demonstrated strong internal consistency and valid measurement capabilities, yet its real-world ecological impact and generalization require further investigation.

These results underscore the importance of aligning intervention design more closely with children's everyday contexts and suggest critical directions for refining digital tools: maintaining effective EF training elements, reconsidering emotional and academic outcome targets, and grounding future designs in ethical, user-centered frameworks that integrate lived experience to enhance meaningful benefit. This study's mixed outcomes provide valuable guidance for developers, educators, and researchers aiming to optimize digital supports for neurodiverse children and emphasize the need to distinguish which design features to retain or adapt versus those that may be less effective.

H1. Executive Functioning (EF) Outcomes

Children receiving the EMPOWER intervention demonstrated selective improvements in executive functioning from pre- to post-test. Specifically, significant gains were observed in planning (CHEXI Planning: $t(42) = 2.46$, $p = .018$) and total working memory (CHEXI Total Working Memory: $t(42) = 2.31$, $p = .026$) within the intervention group, with a trend toward improvement in working memory subscale ($p = .071$). However, other EF domains (inhibition, regulation, attention, flexibility) did not show significant within-group change, and there were no significant group differences at post-test between EMPOWER and TAU (t-tests $ps > .05$). In direct task measures, neither group showed robust improvement in working memory, attention, or cognitive flexibility, possibly due to limited intervention dosage or statistical power. Thus, partial support is found for H1, most strongly for planning and working memory sequencing in the EMPOWER group.

H2. Emotion Regulation (ER) Strategies

Results showed that at post-test, children in the EMPOWER group (compared to TAU) reported significantly less use of maladaptive emotion regulation strategies (AERQ Maladaptive: $t(82)=3.91$, $p=.001$, with lower means for the EMPOWER group) and suppression (AERQ Suppress: $t(82)=3.79$, $p=.001$), both reflecting more adaptive ER. No significant differences were found for other ER subscales (situation selection, redirecting attention, reappraisal, social support, adaptive total). Therefore, H2 is supported in that EMPOWER participants use fewer maladaptive regulation strategies post-intervention, even though broad improvement across all subdomains was not observed.

H3. Emotional Symptoms

No significant differences were detected in emotional and behavioral outcomes (SDQ Emotional Symptoms subscale) between groups at post-test ($t(81.9) = 0.63$, $p = .529$), nor significant pre–post changes within the EMPOWER intervention group ($t(42) = 1.35$, $p = .185$). H3 is not supported; the intervention did not yield a clear reduction in emotional symptoms relative to TAU.

H4. Academic Success and Productivity

Academic performance as measured by the APRS showed no statistically significant post-test group differences in academic success ($t(82) = 2.51$, $p = .014$ in favor of control, but control already had higher pre-intervention scores), nor for productivity or impulse control. No clear intervention effect was observed in the EMPOWER group compared to TAU after accounting for baseline differences. Therefore, the data do not support H4; the EMPOWER intervention did not produce improved academic outcomes compared to TAU in this sample.

H5. EMPOWER Platform Validity and Reliability

The platform demonstrated high internal consistency for game-based and questionnaire-derived EF and ER metrics (e.g., Cronbach's α 0.90–0.96 for core scales), showed robust pre–post improvements in targeted EF domains, and effectively recorded longitudinal progress in digital tasks (notably, attention and processing speed). The multi-method, multi-rater design, convergence of observational and direct metrics, and adaptive intervention features all contribute to its face and content validity. However, limitations include limited generalization to daily life, variable real-world transfer, and the need for further validation of ecological impact and long-term outcomes. EMPOWER can be considered a valid and reliable tool for training and tracking EF and ER in neurodiverse children, but broader generalization remains to be firmly established.

The longitudinal data from the **sustained attention game** demonstrate clear performance gains in a cohort of 42 participants tracked across multiple sessions. Participants responded significantly faster over time, with a mean response time decrease of 0.23 seconds (representing a 13.8% reduction), and accuracy-related metrics improved modestly, as evidenced by increased correct present (CP) responses and decreased error counts. These improvements are accompanied by considerable individual variability, indicating diverse trajectories of attentional skill development among participants.

Statistical analysis across task difficulty levels on the sustained attention game shows that in Level 1, significant gains were observed for error present and correct present scores, indicating increased accuracy and responsiveness in detecting and responding to targets over repeated sessions. The absence of significant improvement in error absent and correct absent conditions suggests that the most pronounced benefits were specific to conditions requiring active detection, rather than to all trial types. Level 2 analyses mirrored this pattern, with significant enhancements in error present and correct present responses, while changes in error absent and correct absent responses did not reach significance. At Level 3, group analysis similarly revealed significant improvement only in correct present and correct absent measures, with error indices unchanged.

These results are consistent with recent findings in cognitive training literature, which show that targeted, adaptive attention games can improve response time and various aspects of attentional control over extended training, often with specificity to the demands emphasized during gameplay. Notably, improvements tend to be most robust for indicators tied to active target monitoring and response execution, skills directly exercised by the game, while generalization to all types of attentional trials is less consistent. The observed reduction in response time alongside increased accuracy points to an overall improvement in processing efficiency. However, individual differences in change trajectories, as revealed by sizeable standard deviations, underline the importance of considering participant variability when assessing intervention effects.

On the **working memory game**, the analyses showed that both groups exhibited significant pre–post improvements in working memory at the lowest difficulty level (Level 1, Correct Order), while no reliable changes emerged at higher task levels. This suggests near-transfer improvements under low cognitive load, but limited effects on more demanding WM operations.

In terms of **cognitive flexibility**, results were most pronounced for Group 1. At Level 1, significant gains were observed for correct responses (CR) and category correct sorting (CCS), without reliable changes in error indices. At Level 2, the outcome pattern was null, likely reflecting limited statistical power due to very small subgroup sizes. At Level 3, the pattern became more complex: CCS improved significantly, while non-perseverative errors increased and overall correct responses declined, with perseverative errors remaining unchanged.

Taken together, these findings suggest that the intervention most consistently enhanced foundational WM sequencing and conceptual/category-based sorting ability. However, at higher CF load, conceptual rule acquisition appeared to improve at the cost of accuracy, indicating possible trade-offs in execution under complex conditions. The results point to selective, load-dependent benefits of the intervention. Improvements in WM processes at Level 1 indicate a strengthening of simple sequencing and maintenance skills, whereas the absence of gains at higher WM levels suggests that the intervention primarily reinforced low-demand processes or was insufficient in intensity to affect higher-level manipulation or the low number of children per group could also influence the results. The catergory sorting improvements at both low and high levels of CF highlight enhanced capacity for rule abstraction and set shifting, which aligns with theoretical accounts emphasizing strategy discovery and rule learning.

Overall, the findings demonstrate specificity rather than broad generalization, consistent with short-term or targeted interventions, where improvements first manifest in foundational processes rather than higher-level domains.

Interpretation of the findings must be tempered by several methodological factors. Some of the analyses were highly underpowered (df = 3), meaning that null findings cannot be taken as evidence of no effect. The large number of statistical tests conducted also increases the risk of Type I error, and more stringent corrections would provide greater confidence in the robustness of the results. In this study, effects were consistently interpreted such that improvements corresponded to increases in accuracy and decreases in error counts. While baseline equivalence strengthens internal validity, analyses such as mixed ANOVAs or difference-in-differences models would offer stronger causal inference regarding intervention effects. Finally, the ecological validity of the gains remains uncertain. It is not yet clear how improvements in simple sequencing or conceptual sorting translate into functional outcomes in academic achievement, behavior, or self-regulation. Future studies should incorporate both proximal measures of strategy deployment and distal measures of academic and daily life functioning to clarify the broader significance of the observed cognitive changes.

3.2. LIMITATIONS

Despite its advanced features, the EMPOWER platform, like other digital solutions for assessing and training executive function and emotion regulation in neurodiverse children, still faces several notable limitations. Many platform tasks are highly structured and gamified, which may not fully capture the complexity and ambiguity of real-world challenges, meaning that children's performance in the platform does not always translate directly to daily functioning outside the digital environment. Additionally, the platform often focuses on a select group of cognitive domains and may not adequately assess contextual control, social cognition, multitasking, or decision-making in real-life situations, potentially overlooking difficulties that arise when children engage in activities requiring multiple executive and emotional skills simultaneously.

Access and equity remain practical concerns, as EMPOWER relies on consistent internet connectivity, compatible devices, and digital literacy among users, which can be unevenly distributed across geographic, socioeconomic, and cultural contexts. Data privacy, the need for specialized supervision, and potential sensory or behavioral responses to wearables and devices further complicate large-scale adoption. Finally, while EMPOWER features AI-driven personalization and progress monitoring, long-term effectiveness for predicting and improving clinical outcomes still requires robust, longitudinal research and validation in diverse populations. These challenges highlight the necessity for ongoing development, real-world integration, and refinement to maximize the impact and accessibility of digital tools like EMPOWER for all neurodiverse children.

#4. References

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on psychological science*, *1*(2), 164-180.

Bausela-Herreras, E., Tirapu-Ustarroz, J., & Cordero-Andres, P. (2019). Executive function deficits and neurodevelopmental disorders in childhood and adolescence. *Revista de neurología*, *69*(11), 461-469. (Ed. impresa). https://doi.org/10.33588/rn.6911.2019133

Berkovits, L. D., Eisenhower, A. S., & Blacher, J. (2014). Emotion dysregulation and social competence: Stability, change and predictive power. Journal of Intellectual Disability Research, 58(8), 745–760. https://doi.org/10.1111/jir.12070

Bombonato, C., Del Lucchese, B., Ruffini, C., Di Lieto, M. C., Brovedani, P., Sgandurra, G., ... & Pecini, C. (2024). Far transfer effects of trainings on executive functions in neurodevelopmental disorders: A systematic review and metanalysis. *Neuropsychology Review*, *34*(1), 98-133. https://doi.org/10.1007/s11065-022-09574-z

Breaux, R., Brown, E. S., Harvey, E. A., & Waschbusch, D. A. (2021). Adolescent emotionality and emotion regulation in the context of parent emotion socialization among adolescents with neurodevelopmental disorders: A call to action with pilot data. Research on Child and Adolescent Psychopathology, 49(11), 1311–1322. https://doi.org/10.1007/s10802-021-00883-4

Burić, I., Sorić, I., & Penezić, Z. (2016). Academic Emotion Regulation Questionnaire (AERQ): Development and validation. Learning and Individual Differences, 52, 119–130. https://doi.org/10.1016/j.lindif.2016.10.014

Cai, R., Leichtman, M., DeRosier, M. E., & Hooper, S. R. (2021). Brief report: Emotion regulation influences on internalizing and externalizing symptoms across the normative-clinical continuum. Frontiers in Psychiatry, 12, 790257. https://doi.org/10.3389/fpsyt.2021.790257

Chiang, H.-L., Gau, S. S.-F. (2014). Impact of executive functions on school and peer functions in youths with ADHD. Research in Developmental Disabilities, 35(4), 963–972. https://doi.org/10.1016/j.ridd.2014.02.010pubmed.ncbi.nlm.nih

Cibrian, F. L., Lakes, K. D., Schuck, S. E. B., & Hayes, G. R. (2021). The potential for emerging technologies to support self-regulation in children with ADHD: A literature review. International Journal of Child-Computer Interaction, 31, 100421. https://doi.org/10.1016/j.ijcci.2021.100421

Costescu, C. A., David, O. A., Roșan, A., Cozma, A., & Pădure, L. (under review). REMEX: Assessing executive functions in children—Direct tasks. Manuscript under review

Diamantopoulou, S., Rydell, A. M., Thorell, L. B., & Bohlin, G. (2007). Impact of executive functioning and symptoms of attention deficit hyperactivity disorder on children's peer relations and school performance. *Developmental neuropsychology*, *32*(1), 521-542. https://doi.org/10.1080/87565640701360981

Diamond, A. (2016). Executive Skills Checklist (ESC). Unpublished measure.

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions that improve executive functions must be reached cautiously. Developmental Cognitive Neuroscience, 18, 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

DuPaul, G. J., Rapport, M. D., & Perriello, L. M. (1991). The Academic Performance Rating Scale. Journal of Psychopathology and Behavioral Assessment, 13(2), 127–137. https://doi.org/10.1007/BF00961235

Galitskaya, V., Polydoros, G., Drigas, A., & Antoniou, A.-S. (2025). Exploring the impact of executive function deficits on academic readiness and social-emotional skills in children with ADHD and learning disabilities: A systematic review. Scientific Electronic Archives, 18(9), xx-xx. https://doi.org/10.36560/18920252078

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. Journal of Child Psychology and Psychiatry, 38(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Hatta, Z. M., & Idris, M. R. (2025). Exploring emotion regulation: Interventions, neurobiological mechanisms, and developmental impacts across children with neurodevelopmental conditions: A systematic review. International Journal of Education, Psychology and Counseling, 10(2), 1–25. https://doi.org/10.35631/IJEPC.102001consensus

Hennessy, A., Nichols, E. S., Al-Saoud, S., Brossard-Racine, M., & Duerden, E. G. (2024). Identifying cognitive profiles in children with neurodevelopmental disorders using online cognitive testing. Clinical Child Psychology and Psychiatry, 29(2), 591–607. https://doi.org/10.1177/13591045241228889

Hernández Lara, M., García-Torres, M., & Benítez, S. (2023). Technology for supporting emotion regulation of individuals with developmental disabilities: A scoping review. Research in Developmental Disabilities, 138, 104503. https://doi.org/10.1016/j.ridd.2023.104503

Howard, S. J., & Melhuish, E. C. (2016). An Early Years Toolbox for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. Journal of Psychoeducational Assessment, 34(3), 255–275. https://doi.org/10.1177/0734282916633009

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. Child Development, 67(2), 490–507. https://doi.org/10.2307/1131828

Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY–II: Clinical and interpretive manual. San Antonio, TX: Pearson.

Kouklari, E. C., Tsermentseli, S., & Monks, C. P. (2018). Everyday executive function and adaptive skills in children and adolescents with autism spectrum disorder: Cross-sectional developmental trajectories. Autism & Developmental Language Impairments, 3, 2396941518800775. https://doi.org/10.1177/2396941518800775

Leonard, H. C., & Hill, E. L. (2015). Executive difficulties in developmental coordination disorder: methodological issues and future directions. *Current Developmental Disorders Reports*, *2*(2), 141-149. https://doi.org/10.1007/s40474-015-0044-8

Li, B., Atyabi, A., Kim, M., Barney, E., Ahn, A. Y., Luo, Y., Aubertine, M., Corrigan, S., St. John, T., Wang, Q., Mademtzi, M., Best, M., & Shic, F. (2018). Social influences on executive functioning in autism: Design of a mobile gaming platform. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Paper 266). ACM. https://doi.org/10.1145/3173574.3174017

Lingala, J., Ram Kumar, B. V., Godishala, S., Rout, D., & Godishala, S. (2021). Executive functions in children with autism: An overview. Elixir Nursing Sciences, 150, 55183–55190

Martin, R. E., & Ochsner, K. N. (2016). The neuroscience of emotion regulation development: Implications for education. Current Opinion in Behavioral Sciences, 10, 142–148. https://doi.org/10.1016/j.cobeha.2016.06.006

Orsoni, M., Giovagnoli, S., Garofalo, S., Magri, S., Benvenuti, M., Mazzoni, E., & Benassi, M. (2023). Preliminary evidence on machine learning approaches for clusterizing students' cognitive profile. *Heliyon*, *9*(3). https://doi.org/10.1016/j.heliyon.2023.e14506

Otterman, D. L., Koopman-Verhoeff, M. E., White, T. J., Tiemeier, H., Bolhuis, K., & Jansen, P. W. (2019). Executive functioning and neurodevelopmental disorders in early childhood: a prospective population-based study. *Child and adolescent psychiatry and mental health*, *13*(1), 38. https://doi.org/10.1186/s13034-019-0299-7

Pardo-Salamanca, A., Paoletti, D., Pastor-Cerezuela, G., Traverso, L., & Luengo, M. A. (2024). Executive functioning profiles in neurodevelopmental disorders: Parent–child outcomes. Children, 11(8), 909. https://doi.org/10.3390/children11080909orcid+1

Ramos-Galarza, C., +3. (2020). Technological innovation to assess cognitive functions in attention deficit hyperactivity disorder. In Proceedings of the International Conference on Applied Human Factors and Ergonomics (pp. xx-xx). AHFE. https://doi.org/10.54941/ahfe1002108

Rosello-Miranda, B., Berenguer-Forner, C., & Miranda-Casas, A. (2018). Adaptive behaviour and learning in children with neurodevelopmental disorders (autism spectrum disorders and attention deficit hyperactivity disorder). Effects of executive functioning. *Revista de neurologia*, *66*(Suplemento 1), 127-132. https://doi.org/10.33588/rn.66S01.2017530

Sadozai, A. K., Sun, C., Demetriou, E. A., Lampit, A., Munro, M., Perry, N., ... & Guastella, A. J. (2024). Executive function in children with neurodevelopmental conditions: a systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2357-2366. https://doi.org/10.1038/s41562-024-02000-9

Sailer, M., & Homner, L. (2020). Gamification in education: A meta-analysis. Educational Psychology Review, 32(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. Computers in Human Behavior, 87, 192–206. https://doi.org/10.1016/j.chb.2018.05.028

Samson, A. C., Huber, O., & Gross, J. J. (2022). An exploratory study on emotion regulation strategy use in individuals with Williams syndrome, autism spectrum disorder and intellectual disability. Frontiers in Psychiatry, 13, 832701. https://doi.org/10.3389/fpsyt.2022.832701

Scaife, M., Rogers, Y., Aldrich, F., & Davies, M. (1997, March). Designing for or designing with? Informant design for interactive learning environments. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (pp. 343-350).

Schultz, D., Izard, C. E., & Bear, G. (2004). The Assessment of Children's Emotion Skills (ACES). Psychological Assessment, 16(4), 328–342. https://doi.org/10.1037/1040-3590.16.4.328

Sun, J., & Buys, N. (2012). Early executive function deficit in preterm children and its association with neurodevelopmental disorders in childhood: a literature review. *International journal of adolescent medicine and health*, *24*(4), 291-299.

Tajik-Parvinchi, D., Wright, L., Frijters, J., Duku, E., & Roberts, S. W. (2021). The role of cognitive control and emotion regulation in predicting mental health problems in children with neurodevelopmental disorders. Child: Care, Health and Development, 47(1), 52–60. https://doi.org/10.1111/cch.12821

Tal-Saban, M., Zaguri-Vittenberg, S., Nevo, Y., Parush, S., & Meyer, S. (2024). The effect of neurodevelopmental disorders characteristics on aspects of social functioning. Physical & Occupational Therapy in Pediatrics, 44(6), 783–797. https://doi.org/10.1080/01942638.2023.2291456pubmed.ncbi.nlm.nih

Thorell, L. B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. Developmental Neuropsychology, 33(4), 536–552. https://doi.org/10.1080/87565640802101516

Vergara, A. I., Balluerka, N., & Gorostiaga, A. (2022). Psychometric properties of children's emotion knowledge and recognition measures: A systematic review with a focus on ACES. Psychological Assessment, 34(10), 893–909. https://doi.org/10.1037/pas0001149

Will, E., Fidler, D. J., Daunhauer, L. A., & Gerlach-McDonald, B. (2017). Executive function and academic achievement in primary-grade students with Down syndrome. Journal of Intellectual Disability Research, 61(2), 181–195. https://doi.org/10.1111/jir.12313onlinelibrary.wiley

Willert, N. (2021). A systematic literature review of gameful feedback in computer science education. Enhancing Learning and Assessment in Computer Science Education, 11(10), 22.

Yusoff, Y. M., Muhammad, Z., Zahari, M. S. M., Pasah, E. S., & Robert, E. (2009). Individual differences, perceived ease of use, and perceived usefulness in the e-library usage. Computer and Information Science, 2(1), 76-83.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational psychologist*, *48*(3), 135-147. https://doi.org/10.1080/00461520.2013.79