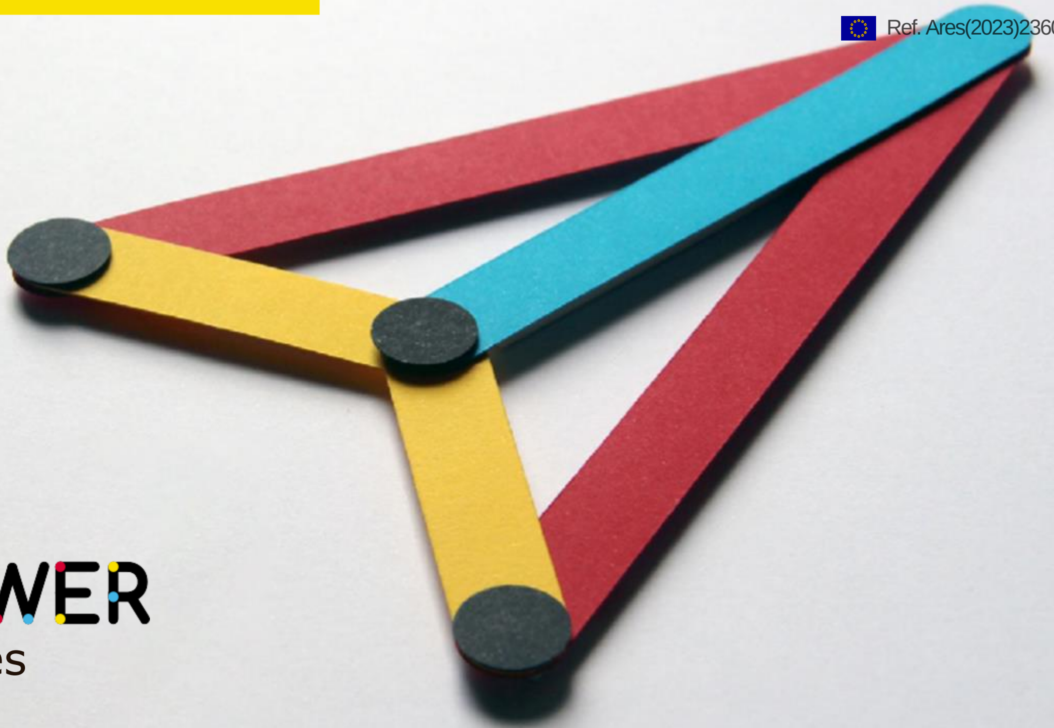
D4.1



EMPOWER
deliverables

Deliverable name — **ALGORITHM FEATURES AND ABILITIES**

Type — **R - Document Report**

Dissemination level — **PU - Public**

Date — **6**

This deliverable documents the features and abilities of the sensory interpretation algorithms and a complete analysis of the requirements, both what is desired by stakeholders such as children, educators, and psychologists (WP2) and what is legally and ethically permissible (WP5).

Description

WP.4

Work Package. 4

Lead Beneficiary – RU

Algorithms Features and Abilities

Executive Summary

This document describes recent work developed by EMPOWER partners involved in WP2, WP3, WP4 and WP5, which aims to combine perspectives on all stakeholders (children, educators and psychologists) and define the scope of algorithms that will be developed in the project.

Date	Version	Description	Authors
09.03.2023	0.1	First Draft: Requirements specification, including constructs definition, standardised tasks definitions, games and in-game metrics.	Cristina Cotescu, Carmen David, Adrian Roșan, Paula Ferreira, Aristides Ferreira, Diana Stillwell, Lucia Vera
24.03.2023	1.0	Creation of high-level view and addition of technical details and consideration	Joana Campos, Marcos Bueno, Serge Thill and Ana Paiva
30.03.2023	1.1	Reviews and edits	Joana Campos, Marcos Bueno, Serge Thill, Ana Paiva, Jerry Lin

Table of Content

#1. Introduction

Children with neurodevelopmental disorders (NDDs[1]) can experience difficulties with language and speech, motor skills, behaviour, memory, learning, or other neurological functions. Moreover, they are 3 to 6 times more likely than their peers to have mental health problems, such as emotional related issues (anxiety or depression) or behavioural problems (Simonoff et. al., 2008), which can lead to impairment in social and academic functioning. Specialised interventions in children with NDDs can help them develop several specific skills or behaviours and can reduce symptoms severity, but children's overall mental health and inclusion rates do not improve significantly. One possible explanation is that 80% of children and adolescents with NDDs present at least a moderate degree of emotional dysregulation (Mayes et al., 2017), meaning that they have difficulties in modulating emotions in the service of one's goals, showing irritable-angry mood and temper outbursts. Standard interventions, however, do not focus on that.

In EMPOWER we aim to develop a game-based (multicultural) learning platform that will help children with NDDs to reduce their emotional and behavioural problems. The new assessment-intervention model, designed to be better integrated into the educational system, aims to leverage a non-intrusive technological approach (biosensors, machine learning, virtual and augmented realities) to augment awareness of the cognitive and emotional development of children with NDDs. Furthermore, effective treatment is not determined by a fixed number of therapeutic sessions but rather by individualised educational plans (Brown, Parikh, & Patel, 2020), therefore the tool will provide **personalised interventions** and **self-adaptation** of the games according to the student's profile, sensors readings and performance within the games' tasks.

This document describes the assessment and adaptations mechanisms that will be developed in the lifetime of the project and which not only do they to take into account the perspectives of the stakeholders (children, educators and psychologists), but also incorporate a developmental and educational perspective by monitoring two important constructs: **Executive Functions** (EFs) and **Emotion Regulation** (ER) strategies.

The assessment and adaptation mechanisms will be supported by Artificial Intelligence (AI) technology, in particular Machine Learning (ML) algorithms, as well as pattern analytic models. ML algorithms build models from historical data, for example, scores and sensory data of students collected in previous sessions, which can be used for making predictions for future situations. Pattern analytic models help to discover insights of the data to identify the relationships of items, e.g, features of the collected data. In this document, we specify the

[1] [1] According to the Diagnosis and Statistical Manual of Mental Disorders (APA, 2013) NDDs comprise intellectual disability, communication disorders, autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), motor disorders and specific learning disorders (SLD). The validity of NDDs as a construct is supported by the high rates of comorbidity between various disorders within this diagnostic grouping (i.e., 22% to 83% of children with ASD have symptoms that satisfy the diagnostic criteria for ADHD, and vice versa, 30% to 65% of children with ADHD have clinically significant symptoms of ASD, Sokolova et.al., 2017).

requirements for such algorithms in detail.

TERMINOLOGY AND ACRONYMS

- **ML**: Machine Learning
- **AI**: Artificial Intelligence
- **ET**: Eye Tracking
- **ML Algorithm:** a process used to create an **ML model** (also called *model training*), normally from historical data. Once trained, a model can make predictions for an outcome variable by supplying data on the remaining variables. We abuse the notation by using the terms algorithm and model interchangeably.
- **Variable, feature**: a quantity such as Age, Gender, Number of Correct Answers, Sequential ordering, Timestamp, etc. We use these terms interchangeably.
- **XAI**: eXplainable AI

DOCUMENT OUTLINE

In the following sections we describe the theory behind the nine key developmental and educational areas of intervention of EMPOWER and the instruments/tasks used to measure or train a skill. Following that we present the pipeline for the SMART training tool developed in this project. We delve into the Algorithms and final notes.

#2. The EMPOWER *Smart Training* Pipeline

This section provides an overview of the EMPOWER Smart Training Pipeline and how the work of the different work packages provides the intended final product. A central aspect of the developed games and thus the work that will be developed in this work package relies greatly on the nine key developmental and educational areas of intervention of EMPOWER. In this section, we define each one of the constructs and present the standardized task used to measure/train the skill. These tasks are the scaffold for the games and how children perform in these tasks is the information necessary for the algorithms. The algorithms should be able to deliver information to the games allowing them to adapt to student performance in real time. The algorithms should also deliver feedback in an adequate form to all relevant stakeholders of EMPOWER, such as teachers, students, and parents.

UNDERLYING PSYCHOLOGICAL CONSTRUCTS

The underlying mechanisms of emotional and behavioural problems of children are classified in relation to impairments in executive functions (EF). Executive functioning is an umbrella term that illustrates a group of high-order cognitive processes that facilitate goal-directed behaviours and an individual's approach in novel situations (Diamond, 2013). In lay terms, it consists of "those capacities that enable a person to engage successfully in independent, purposive, self-serving behaviour" (Lezak et.al., 2012). EF are a complex construct composed of a set of abilities that control and regulate processes such as *attention*, *memory* and *motor skills* (Gilbert et al., 2008).

Assessment of the cognitive and emotional factors plays an important role in the teaching-learning process, and it is a powerful tool for enhancing children's achievement and facilitating their access to personalised interventions (Neumann et al., 2018). In this project, we are interested in the following nine key developmental and educational areas of intervention (further elaborated below), which encompass components of EF:

1) Sustained Attention;
2) Working Memory;
3) Cognitive Flexibility;
4) Delayed Gratification; and
5) Behavioural Inhibition.
6) Emotion naming
7) Emotion Intensity level rating
8) Emotion Understanding
9) Emotion Regulation Strategies

Emotion regulation strategies (and other emotional constructs) are linked to some cognitive processes that are classified as executive functions.

1. ***SUSTAINED ATTENTION*** refers to more than pure attentional focus. Maintenance of focus of attention requires that one maintains in mind what the task is (flexibility to respond to

changes in the environment) and not be susceptible to other disturbances (suppression of irrelevant stimulus). A theoretical position taken by Posner and Petersen (1990) proposed that the attentional system can be broken into three networks represented in distinct anatomical areas: (a) The vigilance network, responsible for maintaining a state of alertness; (b) the visual orienting network, which controls the selection of information from sensory input; and (c) the executive attention network, responsible for resolving conflicts among responses.

2. ***WORKING MEMORY*** refers to a brain system (or systems) necessary to maintain task-relevant information during the execution of complex cognitive tasks, such as reasoning, language, learning and comprehension (Baddeley, 1992). Miyake et al. (2000) identify in their executive functions model, three core components that regulate the dynamics of human cognition: *shifting*, *updating* and *inhibition*. Updating (and monitoring) is closely linked to the notion of Working memory, representing processes that not only do they require maintenance of task-relevant information in memory, but also dynamically manipulate concepts and information. This is how working memory is being addressed in EMPOWER.

3. ***COGNITIVE FLEXIBILITY*** refers to the ability to adaptively switch between different cognitive tasks or mental sets, modifying thoughts and behaviour according to changing task demands (FitzGibbon et al., 2014). According to Miyake et al. (2000), cognitive flexibility *(shifting*) is one of the three core components of executive functioning, alongside working memory and inhibition. Research has demonstrated that cognitive flexibility is essential for several cognitive tasks, including problem-solving, reasoning, and decision-making. In addition, researchers have suggested that cognitive flexibility is associated with better academic performance and social functioning (Diamond, 2013). Importantly, this EF component is conceptualised as a later developing skill resulting from improvements in working memory and inhibition (Blackwell et al., 2014).

4. ***DELAYED GRATIFICATION*** is the extent to which one can resist the temptation of an immediate reward and wait for a larger reward later. It is a self-regulatory skill (Duckworth et al., 2013) that predicts positive outcomes. The ability to delay gratification is one of the skills that may support individuals' self-regulation, affecting how they adapt across many different contexts. Delayed gratification and inhibition are two cognitive processes that are closely related (Rothbart et al., 2004). Studies have shown that individuals with better inhibition skills are more likely to be able to delay gratification (Mischel et al., 1970). Additionally, interventions targeting inhibition have been found to enhance delayed gratification abilities in both children and adults (Diamond & Lee, 2011; Duckworth et al., 2018).

5. ***BEHAVIOURAL INHIBITION*** may be defined as the ability to stop mid-task to regulate behaviour or complete a non-dominant response is supported by independent processes that are both reactive and proactive (Van Hulst et al., 2018; Verbruggen & Logan, 2009). Reactive inhibition measures the speed of the stopping process whereas proactive inhibition, or proactive slowing, involves strategic response slowing in order to complete more challenging tasks while maintaining accuracy (Van Hulst et al., 2018; Verbruggen & Logan, 2009). Finally,

motivational inhibition gauges the avoidance of losses in activities that include feedback or reward contingencies (Cassotti et al., 2014). Interference control is operationalized as the ability to suppress stimuli that may interfere with a response (Cragg, 2016). The fundamental idea is creating agreement (via congruent stimuli [C]) or conflict (via incongruent stimuli [IC]) between values of the target feature (recycle bin colour) and the distractor feature (recycle bin colour/types of trash) when responding to the target feature (Algom et al, 2022). Hence, a Stroop Effect = MRT (IC) − MRT (C). Inhibitory control can predict social-emotional competence. Children who have better inhibitory control abilities tend to have better social skills and less internalising behaviour (Liu et al., 2018).

6. ***EMOTION NAMING / RECOGNITION*** is defined as the ability to identify emotions in oneself and others (Baron-Cohen et al., 1985). Emotion Recognition is an essential part of social development and is considered to be a basic ability that underlies more complex emotional understanding and social skills (Jones et al., 2011).

7. ***EMOTION INTENSITY*** is the magnitude or strength of the experienced or expressed emotion (Frijda et al., 1992, Sonnemans & Frijda, 1994).

8. ***EMOTION UNDERSTANDING*** is the understanding of emotional expressions, internal feelings, and the antecedents and consequences of emotions in self and others (La Bounty et al., 2008).

9. ***EMOTION REGULATION*** is defined by Gross (1998) as “the processes by which individuals influence which emotion they have, when they have them and how they experience and express these emotions” (p275). Even though we categorise emotional strategies as adaptive or maladaptive, these strategies are not generally good or bad, as their adaptability can vary across different contexts (Gross, 1998). Therefore, as Rohlf and Krahne (2015) propose we will define the adaptivity of the strategy in terms of their consequences depending on the situation. For example, we will define the adaptivity of anger in terms of its consequences in aggression and social rejection and the consequence of sadness is usually withdrawal.

Instruments for Measuring and Training the Constructs

Several instruments are available to capture specific components of this complex construct which is Executive Functions. Not only do these instruments have been used to assess executive dysfunctions, but also as a tool to train individuals to improve specific skills. In this section we describe the measuring/training instruments for each of the development and educational areas target in EMPOWER discussed in the previous section. The tasks in each of these instruments are the scaffold of the games developed in EMPOWER.

1. SUSTAINED ATTENTION: Instrument/Task

There are several instruments used to study sustained attention, but the Continuous Performance Test (CPT) is perhaps the most widely used. The CPT task was developed to improve the function of sustained attention (aspect *b)* of Posner’s theory), that is, to be able to maintain the focus of attention on a given task, especially during the presence of distracting stimuli. In children with ASD, an intervention that directly targets attention functions is computerised progressive attentional training (CPAT). CPAT intervention yielded improved academic performance and a decrease in inattention symptoms compared to a matched active control group. The potential benefit of CPAT in children with ASD was recently demonstrated in a pilot study conducted in two primary schools in Birmingham, UK (Spaniol et al., 2021). Their results improved in maths, reading and word coping.

2. **WORKING MEMORY: Instrument/Task**

The Corsi blocks task is a cognitive test used to assess visuospatial working memory. It involves a series of coloured blocks that are presented to the participant in a random order, and the participant must replicate the sequence of blocks by touching or pointing to the blocks in the same order as they were presented. We modified a version of a Corsi block computerized task, as described in Macizo et al. (2016). We added a concurrent task to the spatial-visual task. While processing the spatial location and updating, the children have to sort the peppers on a criterion. This addition we predict that it increases the demands of the task, while not allowing for visual rehearsal strategies or visual fixations..

3. **COGNITIVE FLEXIBILITY: Instrument/Task**

The Wisconsin Card Sorting Test (WCST) (Heaton et al. 1993) is a neuropsychological test used to assess cognitive flexibility and the ability to shift between different mental sets. This instrument is commonly regarded as ''the gold standard executive function task'' (Ozonoff et al. 2005, p. 532). It is a highly sensitive indicator of executive functions, especially such as mental flexibility, planning, and set maintenance.

The task involves presenting the participant with a deck of cards that vary in colour, shape, and number. The participant is asked to sort the cards according to different rules, such as colour, shape, or number, which are not explicitly stated but instead have to be inferred through feedback from the examiner.

4. **DELAYED GRATIFICATION: Instrument/Task**

The Marshmallow Test is the classic paradigm for studying self-control and delay of gratification in both children and adults (Mischel et al., 1970). The experiment was designed to test the ability of children to delay gratification and resist temptation. In the experiment, a child was offered a choice between a small reward (such as a single marshmallow) immediately or a larger reward (such as two marshmallows) if they could wait for a short period of time (usually around 15 minutes) while the researcher stepped out of the room. The child was left alone with the marshmallow and instructed to resist the temptation to eat it until the researcher returned. The EMPOWER game developed to test/train this skill will try to accommodate the characteristics o the Marshmallow Test.

5. **BEHAVIOURAL INHIBITION: Instrument/Task**

Usually, inhibition is measured using the Stroop task, which is a task in which one must name the ink colour of a colour word if there is a mismatch between the ink colour and word. For example, the word GREEN is printed in red ink.

6. **EMOTION NAMING / RECOGNITION: Instrument/Task**

Test of Emotion Comprehension (TEC) (Pons, Harris, & de Rosnay, 2004). The TEC is based on an extensive review of the experimental literature on the development of emotion understanding.

7. **EMOTION INTENSITY: Instrument/Task**

The Emotions Thermometer is a graphic of a traditional analogue thermometer with a scale of emotional intensity. The scaling of emotional intensity was added to a graphic of a thermometer to make it intuitive (Burg, 2004).

8. **EMOTION UNDERSTANDING: Instrument/ Task**

Test of Emotion Comprehension (TEC) (Pons, Harris, & Rosnay, 2004) was constructed on the developmental model proposed by Pons and Harris, 2000. It addresses several dimensions of the emotion understanding construct: recognition of emotions from facial expressions; understanding external causes of emotions; the role of desires in emotions; the role of beliefs in emotions; emotional memory; emotional appearance; mixed emotions; moral emotions. It is adequate for children with ages between 3- 11 years old. It proposes several situations linked to emotions. The children answer by selecting an emotional expression.

9. **EMOTION REGULATION: Instrument/Task**

Rohlf and Krahne (2015) developed and validated a frustrating task designed to elicit anger, based on a tower construction task. The children's task is to build a wooden 10-block tower. A picture of a tower is first presented, and then the children are instructed to build a tower exactly like the one in the picture within 2 min and 40 s in order to receive a reward (a candy in our study). The task is impossible to solve because two blocks are slightly rounded on one side. The authors claim that such an instrument provides more ecologically valid conclusions about one's capacity to regulate anger than self- and parent reports.

SMART TRAINING PLATFORM OVERVIEW

In EMPOWER we propose a new approach to understanding children's strengths and weaknesses in terms of EFs and ER strategies as predictors for their mental health representing a different perspective in terms of intervention and teaching strategies. Namely, by analysing the objective data gained through the platform one can establish the mechanisms which may explain the appearance of challenging behaviours in children with NDDs and develop a personalised intervention based on their psychological profile.

To that end, we have adopted a generic framework (Figure 1) of three interconnected channels (A1: In-Game Metrics, B1: Sensor Data and C: Ground Truth) that enable necessary data acquisition and analysis. It is the ground truth that will help to make sense of the acquired data and help define and describe the measured constructs (defined in the previous section) in operational terms.

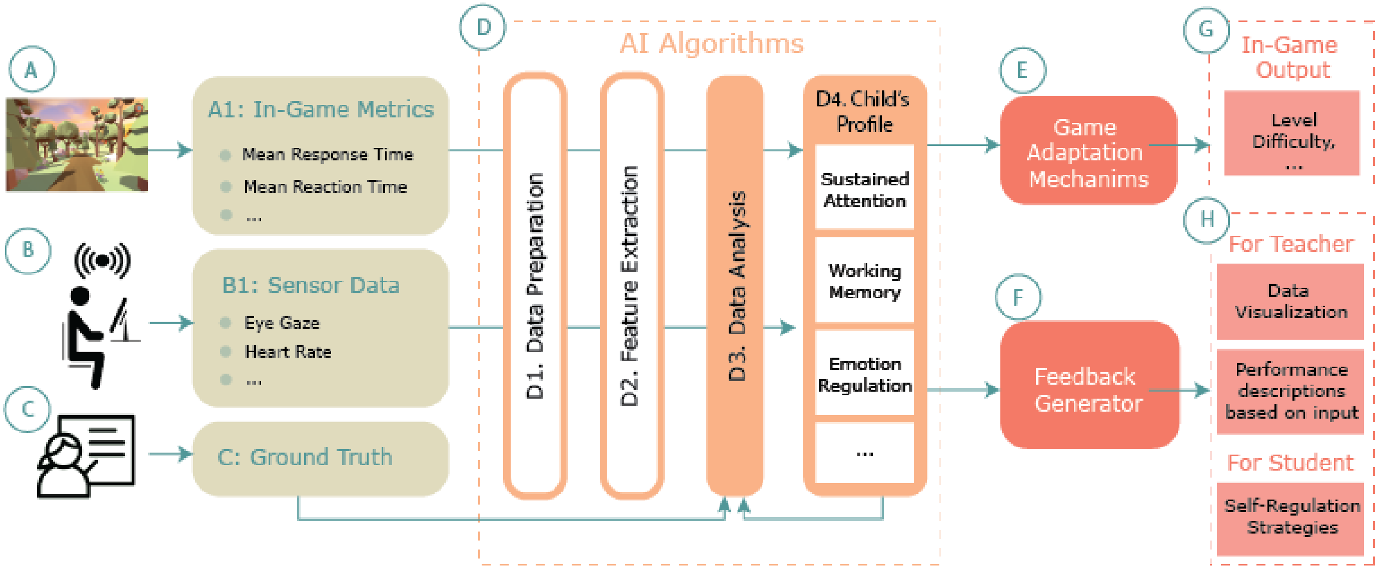


Fig.1 - The Smart Training Pipeline

The components of the conceptual framework can be described as follows:

Games [A]

Games are being designed so that they enable measuring and training the nine key developmental and educational areas of intervention (see Figure 2). At the moment the ECOFARM environment has 3 game prototypes as the starting point in their participatory design: *Attention Game*, *Working Memory Game* and *Inhibition Game* (further described in D3.1). The game allows the assessment of the performance of the user according to the requirements of the standardised task corresponding to the construct.

The ***Attention Game*** is based on a continuous performance test (CPT) and was developed to improve the function of sustained attention, that is, to be able to maintain the focus of attention on a given task, especially during monotonous activities. The task involves a long series of stimuli presented (mostly) sequentially with the participant instructed to respond as fast as possible only when a pre-specified target (e.g., brown mushroom) is presented while withholding responses to other stimuli (e.g., flowers, small branches, butterflies (targets) appearing on the screen).

The ***Working Memory Game*** requires that the users remember which peppers, in a pepper patch, had turned to yellow – in sequence – and which ones are good to sell. This game is a modified version of a Corsi block computerised task. We added a concurrent task to the spatial-visual task. While processing the spatial location and updating, the children have to sort the peppers on a criterion. This addition we predict that it increases the demands of the task, while not allowing for visual rehearsal strategies or visual fixations.

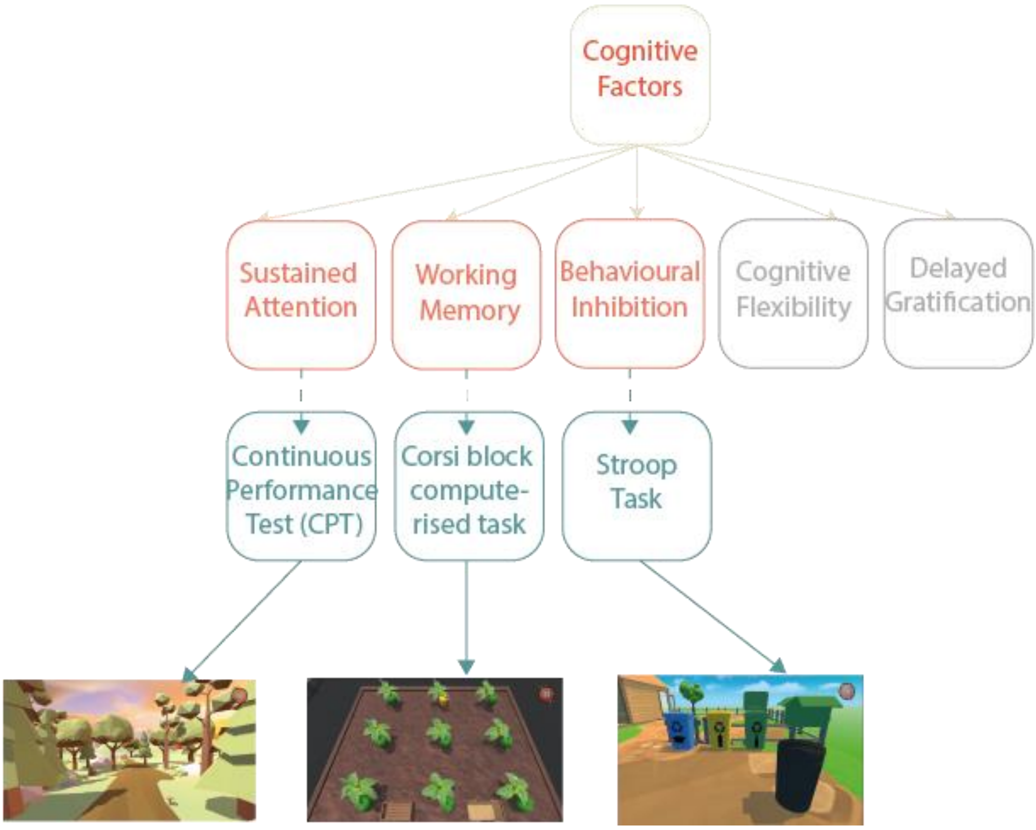


Fig 2. From constructs to games. This figure illustrates the link between theoretical constructs, standardised tasks to measure each of the constructs and the designed game.

In the ***Inhibition Game*** there is a garden full of trash. First, the child has to pick up all the trash in a black bin, and then recycle them in the correct bin. For that, when an item is shown, you have to select if the open bin is correct or not for it. For instance, there if there is a plastic bottle in a green colour coming out of the black bin and the green bin lid is opened the child is confronted with an *incongruent situation* so that the child has to answer “No”. In another case, there can be a yellow box and the blue bin lid opened. In this case, it is a congruent situation so that the child has to answer “Yes”. This game is intended to train the cognitive system to control one’s attention at an optimal level for prolonged periods (as in the Stroop Task) while simultaneously inhibiting the response system.

In-Game Metrics [A1]

Attention Game. These tasks measure if the child is capable of sustaining his/her attention and the measurements derived from the tasks are: (i) The number of correctly identified target stimuli; (ii) The number of omitted target stimuli; (iii) The number of times the participant presses the button when the target is not presented; (iv) The number of times the participant did not press the button when the target is presented.

Working Memory Game. Successful performance in this game is when the child recalls the entire visual-spatial sequence, location and order, while concurrently performing the sorting task based on the indicated criterion. Measurements derived from the task are (i) the longest sequence of recalled items in the right order; (ii) a percentage of total correct trials. A correct trial is one in which the child recalls the yellow peppers in the correct order. It does not depend on the accuracy of the sorting task

Inhibition Game. As in the Stroop Task, we intend to have access to the ratio between congruent and incongruent stimuli. Therefore, measurements derived for this task are (i) the number of correct answers (items in the correct bins); (ii) The reaction time in each trial (assuming the trial was correct); (iii) The mean reaction time of correct.

Physiological Data capture [B]

Assessment of the cognitive and emotional factors plays an important role in the teaching-learning process, and it is a powerful tool for enhancing children's achievement and facilitating their access to personalised interventions (Neumann et al., 2018). Technology such as *eye tracking*, *HR rate sensors*, and *motion sensors* (such as accelerometer, gyroscope and magnetometer) can be found in off-the-shelf wearable devices making easier the adoption and potentially be integrated into educational contexts. Such devices have the potential to deliver better-quality assessments that are more useful for teachers and that more readily benefit students learning (Koomen and Zoenetti, 2018).

Sensors Data [B1]

EMPOWER aims at using the knowledge provided by the wearables as complementary information in the evaluation of NDD children and prediction of potential NDD-specific behaviours. There is some evidence in the literature that eye gaze, HR rate and body kinematics can be variables of interest in adapting an educational activity.

Eye-tracking technologies are commonly used to examine social attention and social motivation in ASD, but their sensitivity varies. Eye-tracking provides insight into the attentional relationships of performance. The ecological significance of social stimuli is an important consideration for measuring social attention and motivation in ASD. The main focus here will be on the analysis of basic, static and dynamic complex emotions based on eye movement parameters such as the *time one looks at an object*, *total fixation duration*, *fixation duration*, *fixation number*, *first fixation*, *saccades* and *smooth pursuit*.

Researchers have found that eye movement predicts working memory performance during the encoding and recall phases (Hodgson, T., 2019). Moreover, when working memory deficits are present, refixations or longer fixation duration are put to use. Using a continuous performance test (CPT), to measure sustained attention, Lev et al. (2020) found that ADHD patients spent more time gazing at irrelevant regions, either on the screen or outside of it. Another study performed by Vakil et al. (2016) demonstrates that adult individuals with ADHD also have a particular pattern of eye gaze (such as total fixation duration, and transitions between target and distractor stimuli) when performing the Stroop task. Overall, research suggests that eye tracking in combination with the actions in the game may enhance the classification accuracy of standard tasks, but further research is necessary to validate the findings for children with NDD.

Heart Rate (HR) is defined as the number of contractions of the heart per minute, and heart rate variability (HRV) which refers to the variations between consecutive heartbeats is widely used to investigate the impact of emotion. HRV is widely used to gather implicit measurements of arousal, although a study conducted by Choi et al. underlines the HRV-based classification of emotion should only be used when a high level of emotion is induced. In EMPOWER measures of HR are important across all possible scenarios as it could be a measure of stress, in particular during an Emotion Regulation Task, during which children will be exposed to a high-level frustration task.

Motion data (kinematics) can provide insightful information about the movements of children in the presence of tasks relevant to stimuli, such as hesitation, coordination between eye and hand movement, and a combination of eye and body movements to compare responses to congruent and incongruent stimuli.

Ground Truth [C]

As in any machine learning pipeline the Ground Truth refers to the information that is known to be true. In EMPOWER the Ground Truth regarding the state of the children in each of the nine key development constructs will be measured by applying questionnaires (about the children) to Teachers. The questionnaires are: **The Childhood Executive Function Inventory** (CHEXI) (Thorell & Nyberg, 2008) – working memory, planning, inhibition, regulation; **Strengths and Difficulties Questionnaire** (SDQ; Goodman, 1997) - emotional and social and peer difficulties; **Academic Emotion Regulation Questionnaire** (AERQ, Buric et al., 2016) - for emotion regulation strategies; **Executive Skills Checklist** (ESC) - ADDitude. Academic performance will also be measured. For more details about the questionnaires and performance measures, refer to Appendix A.

In addition to these, we consider the possibility of having the teacher as a “real-time annotator” while the child is playing the game, as this would give us more accurate data to build the algorithms.

AI Algorithms [D]

The platform will include the measurement of behavioural, subjective, and physiological data of both cognitive and emotional factors. The innovation of this assessment and intervention platform consists of the fact that includes a transdiagnostic approach, measuring the possible underlying mechanism of behavioural and emotional problems of children with NDDs, using direct interaction with the game coupled with data from the physiological sensors. Transdiagnostic treatments are those that specifically target psychological processes or core vulnerabilities that have been observed to contribute to the development and maintenance of a class of disorders, in our case EFs and ER strategies. The sensory interpretation algorithms will process task scores and sensor data in order to provide several functionalities, which are described in the next section. The output of the AI algorithms will feed directly Game Adaptation Mechanisms [E] and Feeback Generator [F] modules. In this component, D1 and D2 refer to offline processes while D3 and D4 to online execution steps.

After the first round of data collection, (including game data and physiological data of the students), it is possible to identify the relationships between individual behaviours and actions (D3). At the end of the data collection, we will be able to understand and evaluate whether the gamified tasks improve the academic performance of the students. Various pattern analytic models such as association rule mining, sequential pattern mining, and utility mining in combination with timestamp and sequence order constraints can be used and applied to find more meaningful and explainable patterns of learning progress. The child’s profile (D4) will encapsulate their performance in each of the constructs and the set of *patterns of behaviours* that are impacting their academic performance.

Game Adaptation Mechanisms [E]

Game Adaptation Mechanisms refer to a set of algorithms that allow to adapt the content of the game to the skills of the child. For instance, verify that the jump between levels is small/large for the individual and make adaptations. In addition, where the child is looking and the speed of stimuli can also trigger adaptation mechanisms. This and other Game Adaptation Mechanisms will be decided after the first pilot study.

Feedback Generator [F]

All the data that is collected about the children during the learning activity, which goes beyond interaction actions and uses eye tracking technology and data from physiological sensors feeds the AI algorithms module.

For the adoption of technology to be successful in education, teachers need to have a deep understanding of how technology can be used to enhance teaching and learning, and how to effectively integrate it into their practice. Among other factors, teachers need to know how technology can inform decision-making. This means that the SMART training platform needs to gather formative and summative feedback on student learning. Feedback will be provided to students and teachers.

In-Game Output [G]

In-Game Output refers to the features in the game that are mutable and susceptible to changes imposed by the game adaptation mechanisms module.

Teacher, Parents and Student Feedback [H]

Student feedback will be given through the self-report of self-regulated learning processes. Specifically, children will be able to register forethought, execution and monitoring, as well as self-reflection through specific target oral questions with their own avatars during specific moments of the session. Specifically, forethought will be reported before the gameplay of each game, but after instructions and example items (e.g., motivation, intention, self-efficacy and learning outcome expectations). Also, self-monitoring during execution will be reported by children (e.g., motivation and metacognition) after a trial and teachers will co-regulate with the child by also reporting observed behaviour. Lastly, children will report their self-reflection about the games at the end of each one (e.g., task value, perceived difficulty, self-assessment of performance).

It is important for machine learning to be explainable in education because it enables teachers and students to understand how the algorithm arrived at a particular suggestion and to trust and interpret the results with confidence. ***Feedback to teachers and parents*** will be provided through natural language explanations for the model's output, in addition to some visualizations. This enables them to understand the reasoning behind the algorithm's suggestions in a more intuitive and human-readable way.

#3. Algorithm Abilities

The general purpose of the algorithms is to support decision-making in different situations in EMPOWER. This includes, e.g., supporting game adaptation, providing recommendations involving the gamified EF and ER tasks, as well as giving insights to teachers about student performance.

The algorithms first create a ML model from data already collected in the platform, e.g. historical performance on tasks and sensor measurements of multiple students. This process is known as *model training*. After training a predictive model, it can be used for making predictions for new cases, e.g. predict the performance of a new student based on this student's sensor measurements. Typically, model training is performed only once and in an *offline* way, as opposed to using the model (normally *online)*. Exceptions exist, e.g., in the situation of data streams and concept drifts when the model can be *retrained* to adapt to new data.

We describe the goals and challenges for the algorithms in EMPOWER that we elicited together with other project partners. We start off by describing the data that the algorithms expect to process.

SOURCES OF DATA

In-game data

As discussed in the previous section, when a student plays a game, data is collected about the student's performance on the specific EF or ER factor that the game targets. This includes multiple task items, from which a final performance score is calculated (see the appendix for further detail). This is also referred to as in-game metrics in Fig-1-A1, or task scores.

Sensor data

The algorithms will also use data collected from sensors such as eye tracking (ET) and smartwatches. This is because sensory data can be linked to affective factors, e.g. student arousal, lack of focus, etc, which will allow the algorithms to have a more complete picture of the student in real-time.

Ground truth data

Refers to student assessment by teachers, e.g. before and after an intervention. This is illustrated by the ground truth block in Fig. 1-C. One example is the emotional and cognitive assessment using questionnaires such as the Strengths and Difficulties Questionnaire (see the previous section for further detail).

STATIC AND TEMPORAL DATA

The types of data described above can be *static* if no sequential or temporal aspect is captured, or *temporal/sequential* otherwise. For example, if a student plays multiple games (even the same game), that is the case of temporal data. Sensor data is likely also temporal since it captures the dynamics of (repeated) bodily measurements over a period of time. Questionnaires can be temporal data if they are completed more than once for a student. It is important to collect the *right data* in the *right way* so that the algorithms can be effective in EMPOWER.

In practice, the algorithms will obtain in-game data from the game engine; sensor data will be obtained from sensor devices, e.g. ET devices; and ground truth data (e.g. questionnaires) will be properly obtained from the teachers.

As agreed with project partners, the algorithms expect that sensor data (ET and smartwatches) and in-game metrics are delivered already preprocessed. For the case of sensor data, the relevant features derived from the sensors should already be computed, such as fixations, saccades, etc. Moreover, all this data shall be delivered to the algorithms in an established format for ML algorithms, such as CSV (comma-separated values) files.

OVERVIEW OF FUNCTIONALITIES

Based on the work developed in WP2, and WP3 and the overall requirements of EMPOWER, we identified several goals for the algorithms. We categorise them based on their primary intended usage.

Recommendations

Information was given by the algorithms to help teachers in decision-making.

1. What is the best starting EF/ER task for a given student?
2. What is the best task/game for a student to play next, given the tasks the student has already played?

Understanding

Insights from the algorithms to a broader range of stakeholders (researchers, teachers, etc.) and external research community. Although there is literature covering some of these topics, it is still unexplored in the broad context of EMPOWER.

3. Are there links between different EFs and ER tasks, e.g. based on student performance in the tasks?
4. Which EFs predict academic performance (or another outcome, such as SDQ) best?

Technical challenges

Challenges to be tackled for the effective functioning of the algorithms in EMPOWER.

5. *Game adaptation*: support game adaptation as needed. The game adaptation details will be known after the pilot study.
6. *Learner model*: what is the best representation of a student in the AI system? An example would be a learner model, where internal aspects could be modelled.
7. *Support for data streams*: Should the algorithms be retrained on a regular basis, e.g. when new labelled/annotated data is generated in the platform? Is concept drift expected to occur?
8. *Robustness indication:* The algorithms should indicate the uncertainty next to its predictions, e.g. in the form of a probability or a statistical interval.
9. *Multi-modal data*: What are the challenges of integrating multi-modal data, i.e. ET, smartwatches, etc?

Feedback

Feedback from/to students and teachers is an essential part of EMPOWER.

10. *Feedback by students*: How can feedback given by students, such as self-report and self-regulated learning, be used for improving the AI system?
11. *Feedback to students*: children will be able to register forethought and self-reflection through specific target oral questions with their own avatars during specific moments of the session.
12. *Feedback from teachers*: How can feedback from the teachers be integrated into the AI system, e.g. correcting an algorithmic recommendation (prediction)?
13. *Feedback to teachers*: Insights about student performance, visualisations, algorithmic explanations (eXplainable AI, XAI), etc. to support teachers in decision-making, e.g. when to make a switch between games. The results can be expected from the XAI or based on the pattern analytic models to analyze the relationships of behaviours and activities of the collected data.

ETHICAL CONSIDERATIONS

The core ethical considerations for AI in EMPOWER will be fully developed within WP5, whose first deliverable will be issued in month 12. Despite that, we already anticipate some ethical considerations for the algorithm design.

We will favour *white-box* models, such as Bayesian networks and tree-models like decision trees and random forests because they are amongst the most *interpretable* AI algorithms. Scientific evidence indicates that less interpretable (known as *black-box* models) such as deep neural networks can be more accurate. However, such black-box models very often need large amounts of data (i.e. a number of students and sessions) to be effectively trained. This data abundance is not the case in EMPOWER.

By using white-box models, the AI systems to be deployed in EMPOWER offer more *transparency*, which is a core ethical aspect. In addition, the models we plan to use are natively more *understandable*, which means that even without using modern XAI techniques one can already gain insight by simply inspecting the trained model.

Finally, we also plan to mitigate potential biases and fairness issues not only at the model level, but also already from the start of the data collection procedure.

#4. Conclusions and Future Steps

This deliverable outlines the desiderata for the SMART training tool of EMPOWER, providing a framework for the game suite that focuses on developing Executive Functions and Emotional Regulation skills. The nine key development constructs define the skills to be trained and are the scaffold for the games. The data of the games will be augmented with data from the sensor readings to address the main needs of the stakeholders. These include identifying the executive functions that best predict emotional and social problems; exploring inter-game performance relationships; and determining how game behaviour predicts standardized task scores. In addition, we recognize the importance of presenting the data in a user-friendly format that teachers, students, and parents can trust and adapt to their needs.

The next steps in the following months will include incorporating students' and teachers' feedback on the design of the games and verifying whether the games are measuring what they are supposed to measure while establishing the ground truth.

References

Algom, D., Fitousi, D., & Chajut, E. (2022). Can the Stroop effect serve as the gold standard of conflict monitoring and control? A conceptual critique. Memory & Cognition, 50(5), 883-897.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37–46

Brown, K. A., Parikh, S., & Patel, D. R. (2020). Understanding basic concepts of developmental diagnosis in children. *Translational pediatrics*, *9*(Suppl 1), S9–S22.

Cassotti M, Ania A, Osmont A, Houdé O, & Borst G (2014). What have we learned about the processes involved in the Iowa Gambling Task from developmental studies? Frontiers in Psychology, 5, 1–5.

Choi, K. H., Kim, J., Kwon, O. S., Kim, M. J., Ryu, Y. H., & Park, J. E. (2017). Is heart rate variability (HRV) an adequate tool for evaluating human emotions?–A focus on the use of the International Affective Picture System (IAPS). *Psychiatry Research*, *251*, 192-196.

Cragg L (2016). The development of audio-visual stimulus and response interference. Developmental Psychology, 52(2), 104.

Diamond A. Executive functions. Annu Rev Psychol. 2013;64:135-68.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959-964.

Duckworth AL, Tsukayama E, Kirby TA. Is it really self-control? Examining the predictive power of the delay of gratification task. Pers Soc Psychol Bull. 2013 Jul;39(7):843-55.

Duckworth, A. L., Milkman, K. L., & Laibson, D. (2018). Beyond Willpower: Strategies for Reducing Failures of Self-Control. *Psychological science in the public interest : a journal of the American Psychological Society*, *19*(3), 102–129.

FitzGibbon, L., Cragg, L., & Carroll, D. J. (2014). Primed to be inflexible: The influence of set size on cognitive flexibility during childhood. *Frontiers in Psychology*

Frijda, Nico H., Andrew Ortony, Joep Sonnemans, and Gerald L. Clore (1992), "The Complexity of Intensity: Issues Concerning the Structure of Emotion Intensity," in Emotion, ed. Margaret S. Clark, Newbury Park, CA: Sage, 60-89.

Sonnemans, Joep and Nico H. Frijda (1994), "The Structure of Subjective Emotional Intensity," Cognition and Emotion, 8 (July), 329-350.

Gilbert, S. J., & Burgess, P. W. (2008). Executive function. *Current Biology*, *18*(3)

Gross, J. J. (2001). Emotion regulation in adulthood: Timing is everything. *Current directions in psychological science*, *10*(6), 214-219.

Heaton, R. K., & Staff, P. A. R. (1993). Wisconsin card sorting test: computer version 2. *Odessa: Psychological Assessment Resources*, *4*, 1-4.

Hodgson, T., Ezard, G., and Hermens, F. (2019). Eye movements in neuropsychological tasks. Current Topics in Behavioral Neuroscience.

LaBounty J., Wellman H. M., Olson S., Lagattuta K., Liu D. (2008). Mothers' and fathers' use of internal state talk with their young children. *Social Development*, 17, 757–775.

Lev, A., Braw, Y., Elbaum, T., Wagner, M., & Rassovsky, Y. (2022). Eye tracking during a continuous performance test: Utility for assessing ADHD patients. *Journal of Attention Disorders*, *26*(2), 245-255.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2012). *Neuropsychological assessment (5th Edition)*. Oxford University Press, USA

Liu, R., Calkins, S. D., and Bell, M. A. (2018). Fearful inhibition, inhibitory control, and maternal negative behaviors during toddlerhood predict internalizing problems at age 6. *J. Abnorm. Child Psychol.* 46, 1665–1675.

Macizo, P. , Soriano, M. F., Paredes, N. (2016). Phonological and visuospatial working memory in Autism spectrum disorders. Journal of Autism and Developmental Disorders, 46, 2956- 2967.

Mayes, S. D., Calhoun, S. L., Waxmonsky, J. G., Kokotovich, C., Baweja, R., Lockridge, R., & Bixler, E. O. (2019). Demographic Differences in Disruptive Mood Dysregulation Disorder Symptoms in ADHD, Autism, and General Population Samples. *Journal of attention disorders*, *23*(8), 849–858.

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of personality and social psychology*, *16*(2), 329.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol. 2000 Aug;41(1):49-100.

Ozonoff, S., South, M., & Provencal, S. (2005). Executive functions. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders:*

Diagnosis, development, neurobiology, and behavior (pp. 606–627). John Wiley & Sons, Inc..

Pons, F., Harris, P. L., & De Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European journal of developmental psychology*, *1*(2), 127-152.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25-42.

Rohlf, H. L., & Krahé, B. (2015). Assessing anger regulation in middle childhood: Development and validation of a behavioral observation measure. *Frontiers in psychology*, *6*, 453.

Rothbart, M. K., Ellis, L. K., & Posner, M. I. (2004). Temperament and self-regulation. *Handbook of self-regulation: Research, theory, and applications*, *2*, 441-460.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(8), 921–929.

Sokolova, E., Oerlemans, A.M., Rommelse, N.N. *et al.* A Causal and Mediation Analysis of the Comorbidity Between Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). *J Autism Dev Disord* **47**, 1595–1604 (2017).

Spaniol, M. M., Mevorach, C., Shalev, L., Teixeira, M. C. T., Lowenthal, R., & de Paula, C. S. (2021). Attention training in children with autism spectrum disorder improves academic performance: A double-blind pilot application of the computerized progressive attentional training program. *Autism Research*, *14*(8), 1769-1776.

Vakil, E., Mass, M., & Schiff, R. (2019). Eye movement performance on the stroop test in adults with ADHD. *Journal of attention disorders*, *23*(10), 1160-1169.

Van Hulst BM, De Zeeuw P, Vlaskamp C, Rijks Y, Zandbelt BB, & Durston S (2018). Children with ADHD symptoms show deficits in reactive but not proactive inhibition, irrespective of their formal diagnosis. Psychological Medicine, 48(15), 2508–2514.

Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience and biobehavioral reviews*, *33*(5), 647–661.

#A. Appendix

The Childhood Executive Function Inventory (CHEXI) (Thorell & Nyberg, 2008) – working memory, planning, inhibition, regulation

Fill in the total score for the respective subscales in the boxes below, and the total score for the two factors WORKING MEMORY and INHIBITION. For an example of ADHD and control group means and SDs, as well as cut off scores, see Catale, Meulemans, & Thorell (in press)[1].

= Subscale 1: Working memory
Total score for items: 1, 3, 6, 7, 9, 19, 21, 23, 24

= Subscale 2: Planning
Total score for items: 12, 14, 17, 20

WORKING MEMORY
Total score

= Subscale 3: Regulation
Total score for items: 2, 4, 8, 11, 15

= Subscale 4: Inhibition
Total score for items: 5, 10, 13, 16, 18, 22

INHIBITION
Total score

The CHEXI includes four different subscales tapping working memory, planning, regulation and inhibition. However, factor analysis in children in kindergarten was only able to identify two factors referred to as **WORKING MEMORY** (working memory and planning subscales) and **INHIBITION** (regulation and inhibition subscales).

Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) - emotional and social and peer dificulties

Table 3: Categorising SDQ scores for 4-17 year olds

	Original three-band categorisation			Newer four-band categorisation			
	Normal	Borderline	Abnormal	Close to average	Slightly raised (/slightly lowered)	High (/Low)	Very high (very low)
Parent completed SDQ							
Total difficulties score	0-13	14-16	17-40	0-13	14-16	17-19	20-40
Emotional problems score	0-3	4	5-10	0-3	4	5-6	7-10
Conduct problems score	0-2	3	4-10	0-2	3	4-5	6-10
Hyperactivity score	0-5	6	7-10	0-5	6-7	8	9-10
Peer problems score	0-2	3	4-10	0-2	3	4	5-10
Prosocial score	6-10	5	0-4	8-10	7	6	0-5
Impact score	0	1	2-10	0	1	2	3-10
Teacher completed SDQ							
Total difficulties score	0-11	12-15	16-40	0-11	12-15	16-18	19-40
Emotional problems score	0-4	5	6-10	0-3	4	5	6-10
Conduct problems score	0-2	3	4-10	0-2	3	4	5-10
Hyperactivity score	0-5	6	7-10	0-5	6-7	8	9-10
Peer problems score	0-3	4	5-10	0-2	3-4	5	6-10
Prosocial score	6-10	5	0-4	6-10	5	4	0-3
Impact score	0	1	2-10	0	1	2	3-10
Self-completed SDQ							
Total difficulties score	0-15	16-19	20-40	0-14	15-17	18-19	20-40
Emotional problems score	0-5	6	7-10	0-4	5	6	7-10
Conduct problems score	0-3	4	5-10	0-3	4	5	6-10
Hyperactivity score	0-5	6	7-10	0-5	6	7	8-10
Peer problems score	0-3	4-5	6-10	0-2	3	4	5-10
Prosocial score	6-10	5	0-4	7-10	6	5	0-4
Impact score	0	1	2-10	0	1	2	3-10

Note that both these systems only provide a rough-and-ready way of screening for disorders; combining information from SDQ symptom and impact scores from multiple informants is better, but still far from perfect.

Executive Skills Checklist (ESC) - ADDitude

Response Inhibition
4 - 20 points

Working memory
10 - 50 points

Emotional Control
4 - 20 points

Sustained attention
3-15 points

Flexibility
4-20 points