American Journal on Intellectual and Developmental Disabilities

Modeling the latent factor structure of the “EXcEEDS” (EX ecutive function Early Evaluation in Down Syndrome) Battery

--Manuscript Draft--

Manuscript Number:	AJIDD-D-24-00046R1
Article Type:	Research Report
Keywords:	Executive Function, Down Syndrome, Psychometric evaluation, Outcome measures
Corresponding Author:	Deborah Fidler Colorado State University Fort Collins, Colorado UNITED STATES
First Author:	Kaylyn Van Deusen
Order of Authors:	Kaylyn Van Deusen
	Mark A Prince
	Madison Walsh
	Lina Patel
	Miranda Pinks
	Anna Esbensen
	Angela Thurman
	Leonard Abbeduto
	Courtney Oser
	Lisa Daunhauer
	Deborah Fidler
Manuscript Region of Origin:	UNITED STATES
Abstract:	Executive function (EF) is frequently an area of vulnerability in conditions associated with intellectual disability, like Down syndrome (DS). However, current EF evaluation approaches are not designed for children with underlying neurodevelopmental conditions and may not demonstrate construct validity due to interpretational confounds. The current study evaluated the construct validity of a novel EF battery designed to reduce measurement challenges in the assessment of young children with DS. Participants were 124 children with DS (2 to 8 years) who completed a set of adapted EF tasks. Exploratory graph analysis demonstrated that a 2-factor solution (an Inhibition factor and a Working Memory/Flexibility factor) was the best fit for the data, providing evidence of construct validity for the adapted EF battery.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Manuscript
Click here to access/download;Manuscript;Van Deusen et al EXcEEDS Battery Evaluation_Resubmitted.docx

Modeling the latent factor structure of the "*EXcEEDS*" (EXecutive function Early Evaluation in Down Syndrome) *Battery*

Abstract

Executive function (EF) is frequently an area of vulnerability in conditions associated with intellectual disability, like Down syndrome (DS). However, current EF evaluation approaches are not designed for children with underlying neurodevelopmental conditions and may not demonstrate construct validity due to interpretational confounds. The current study evaluated the construct validity of a novel battery designed to reduce measurement confounds in the assessment of EF in young children with DS. Participants were 124 children with DS (2 to 8 years) who completed a set of adapted EF tasks. Exploratory graph analysis demonstrated that a 2-factor solution (an Inhibition factor and a Working Memory/Flexibility factor) was the best fit for the data, providing evidence of construct validity for the adapted EF battery.

Modeling the latent factor structure of the *EXcEEDS* (EXecutive function Early Evaluation in Down Syndrome) *Battery*

Executive functions (EF) are the cognitive skills used to direct behavior and achieve goals across the lifespan (Müller & Kerns, 2015). EF skills are associated with a range of functional outcomes throughout development in the general population, including academic achievement and occupational function (Barkley & Murphy, 2010; Best & Miller, 2010). EF skills have increasingly become a target for education and intervention planning for children with intellectual disability-related conditions, like Down syndrome (DS). Identifying effective interventions, however, requires valid EF outcome measures. A challenge, particularly in early childhood, is that few EF direct assessment measures have been psychometrically evaluated for use with DS (Pinks et al., 2023; Van Deusen et al., 2023; Walsh et al., 2023), and no previous work has evaluated the construct validity of a collection of such measures.

Executive Function in Down syndrome

Many young children with DS can potentially benefit from advances in EF treatment science. DS is associated with EF challenges throughout all phases of the lifespan (Loveall et al., 2017; Tungate & Conners, 2021). EF component processes are foundational for adaptive skills and academic achievement in children with DS, and employment outcomes in adults with DS (Tomaszewski et al., 2018; Will et al., 2017). Moreover, early aspects of regulation during infancy have been recently linked with EF outcomes during childhood in DS (Fidler et al., 2023). This suggests that evidence-based EF intervention may be a potentially advantageous approach to strengthening early foundations in DS, with the possibility of impactful downstream effects throughout development across functional, academic, and community contexts.

Measuring the treatment effects of novel EF interventions in DS, however, can be challenging. EF laboratory tasks are often designed in ways that introduce measurement confounds (Willoughby & Hudson, 2021), referred to as the ‘task impurity’ issue. Direct assessment measures of EF frequently require the recruitment of additional cognitive skills (e.g., spatial processing) or several EF component processes when aiming to capture performance in a particular component of EF, which can limit the interpretability of EF performances.

Additional complexity is introduced when evaluating EF in children with neurodevelopmental conditions like DS, wherein many domains of development may be broadly impacted, such as language, motor, and self-regulatory skills. EF scores may demonstrate floor effects or have performance moderated by developmental confounds interacting with assessment type (e.g., verbal versus nonverbal response modalities; Tungate & Conners, 2021). The design of currently available EF assessments, which often include language and motor demands, therefore, may compromise their validity when administered to children with DS and other neurodevelopmental conditions (Channell et al., 2021). The valid assessment of EF in DS requires minimizing task-related demands in these areas to reduce potential performance confounds as much as possible.

In addition to reducing task confounds, measures of EF designed for children with neurodevelopmental conditions must be designed to capture performances at younger chronological ages and developmental levels. Investigations of EF in DS have frequently focused on children 6 years or older (Hessl et al., 2016; Lanfranchi et al., 2010; Schworer et al., 2023; Yang et al., 2014). When younger participants are included in EF studies, they often score at the floor of the administered measures (Hessl et al., 2016; Schworer et al., 2023). As a result, it is difficult to accurately characterize early cognitive regulatory function, a key area of developmental vulnerability, in a population that could benefit from interventions that strengthen EF foundations in early childhood. Without appropriate measurement tools at younger ages, it is also impossible to determine whether such early EF interventions are effective when they are implemented. Thus, novel approaches to EF evaluation are needed that sensitively measure performances at more foundational levels, account for a younger overall developmental status, and are delivered in developmentally appropriate, interactive ways.

EF Architecture

Most models of EF involve three core constructs: inhibitory control (the ability to resist prepotent responses), working memory (the ability to temporarily store and manipulate information), and cognitive flexibility (the ability to switch from one set of thoughts to another; Miyake et al., 2000; Müller & Kerns, 2015). To confirm this structure, Miyake et al.'s (2000) influential study used a latent variable modeling approach to examine whether a collection of measures that were designed to assess various EF component processes indeed measured the specific underlying constructs of interest. Using confirmatory factor analysis (CFA) to model performances of neurotypical college students, they found that a three-factor model was an excellent fit for their data, indicating the dissociability of EF component processes and that the three EF component processes are moderately related to one another (Miyake et al., 2000). Their study also demonstrated the utility of a factor structure approach for determining the construct validity of EF measures (Miyake et al., 2000).

Another important dimension in capturing the underlying factor structure of EF performances involves early developmental emergence. Models later in development tend to demonstrate dissociability of EF component processes (working memory, inhibition, cognitive flexibility), but models of early EF performances tend to support a more unitary construct that may serve as an antecedent to later dissociability (Wiebe et al., 2011). Recent fMRI studies demonstrate that the brain networks that support EF are formed by middle childhood (Engelhardt et al., 2019), providing additional context for the progressive dissociability of component processes over time. Yet, the progressive dissociability account of EF raises questions regarding its measurement and factor structure in populations that demonstrate developmental delays and cognitive challenges. It has long been speculated that specific EF component processes may be differentially impacted within various neurodevelopmental conditions (Ozonoff & Jensen, 1999). Working memory, for example, is an area of particular vulnerability for individuals with DS throughout the lifespan (Conners et al., 2011; Loveall et al., 2017; Tungate & Conners, 2021), whereas cognitive flexibility appears to emerge as an area of distinct vulnerability in adolescence and adulthood (Loveall et al., 2017). However, it is unclear whether the architecture of EF in children with DS at young chronological and developmental ages would conform to a unitary or a dissociable underlying structure.

Current Study

The present study was designed to fill an important gap related to the measurement of EF in individuals with DS (Tungate & Conners, 2021) through the evaluation of the factor structure of a novel EF battery designed for young children with DS. The "EXcEEDS (EXecutive function Early Evaluation in Down Syndrome) Battery" includes adapted versions of existing measures of working memory, inhibition, and cognitive flexibility that have been published in the early childhood EF literature and adapted to eliminate expressive language demands and reduce receptive language and motor demands. Tasks were presented in a game-based context that was socially motivating and fun for participants to assess EF factors. Scoring procedures were implemented that allowed for the inclusion of participants demonstrating emerging abilities while performing at the floor of measures, which addresses the critical factor of heterogeneity in the overall degree of delay observed in this population (Channell et al., 2021; Tungate & Conners, 2021). The battery was administered to children with DS between the ages of 2 to 8 years, and exploratory graph analysis (Epskamp et al., 2018) was used to identify the latent factor structure of the EXcEEDS battery measures. Results of this study have implications for advancing early EF assessment in DS, as well as informing the underlying EF architecture during early childhood in this population.

Methods

Procedures

Children with DS were recruited nationally in the U.S. as part of two larger studies. One study was a large multisite investigation open to children with DS 2 to 8 years of age at several research sites in the US. The second study was a longitudinal investigation that invited a previous cohort to participate in an early childhood follow-up between ages 3 to 7 years.

Caregivers were required to have English understanding for informed consent and completion of project procedures. Assessment visits were scheduled for at least six weeks following a child's surgery for heart defects, and visits were also re-scheduled if experiencing otitis media. Anticipated medication changes delayed starting participation for any individual child and children could not have uncorrected hearing or visual impairment. Children were encouraged to wear corrective devices at the assessment sessions but continued in the assessment if they removed these devices (e.g., took off glasses).

Assessment visits overlapped with the COVID-19 pandemic. Approximately 73% of participants were evaluated with precautions to prevent the spread of COVID-19, including face masks, face shields, and scrubs. Families had the opportunity to delay starting the project if they were not comfortable with in-person visits throughout the COVID-19 pandemic. Informed consent was obtained prior to participation, and all research was approved by institutional review boards. Study data were collected and managed using REDCap electronic data capture hosted at {*withheld for review*} (Harris et al., 2009). Participants were offered breaks as needed and were introduced to the tasks in counterbalanced orders.

Participants

Participants were 124 children with DS with a mean age of 5.26 years (standard deviation (SD) = 1.59; Range = 2.47 – 8.71). The participants were 52.4% male, and 86.2% White. Mental age (MA) estimates were calculated from the Stanford Binet 5^{th} Edition Abbreviated IQ (SB5-ABIQ) and the Bayley Scales of Infant and Toddler Development 4^{th} Edition Cognitive Scale (Bayley-4) and participants demonstrated a mean MA of 2.37 years (SD = 0.91; Range = .42 – 6.33). MA estimates were derived from SB5-ABIQ performances for the majority of participants (72.6% of participants; n = 90). MA was derived from Bayley-4 scores in a subset of participants (21.0% of participants; n = 26) to extend the range of MA scores downward for those participants who had MA estimates at the floor of the SB-5, and who had completed both assessments (see Pinks et al., 2023; Van Deusen et al., 2023; Walsh et al., 2023 for substitution procedure). MA values were missing for 6.4% (n = 8) of participants. See Table 1 for additional demographic information.

Measures

"EXcEEDS" Executive Function Battery

Overview. The "EXcEEDS Battery" tasks were designed to capture specific subcomponents of EF per the Miyake et al. (2000) model. The battery included measures of working memory (WM) and the related construct of short-term memory (STM), measures of inhibitory control (Inhib), and measures of cognitive flexibility (Flex). Tasks were selected to reduce demands outside of the target construct, including eliminating requirements to respond with expressive language, reducing receptive language and motor demands, and presenting each task in a game-like fashion. STM and WM were both included because performances are thought to be overlapping at younger developmental ages (Bell et al., 2022) and have previously been included together as indicators of EF in a meta-analysis of individuals with DS (Tungate & Conners, 2021). WM requires the updating and manipulation of information, whereas STM is the temporary storage and retrieval of information. Both have been noted areas of challenge for individuals with DS (Conners et al., 2011), and were both considered important for EF skill acquisition for the chronological age (CA) range in the present study. The "Inhib" measures included in the EXcEEDS Battery each measured behavioral inhibition, which involves resisting a prepotent response. Each of the "Flex" measures in this investigation included the introduction of an initial sorting rule, followed by a request to sort by a "silly" rule that reversed the rule children had been sorting by. The receptive and cognitive demands of switching increased in difficulty across the three measures to capture greater variability in emerging cognitive flexibility skill acquisition.

Scoring Adaptations. The present investigation of the "EXcEEDS Battery" included participants with a wide CA range in order to evaluate the psychometric soundness of measures designed to evaluate emerging EF during early childhood in DS. This necessitated the use of expanded scoring procedures to characterize the nuances of emerging skills even when performances were at the floor of various tasks. Rather than excluding participants at the floor, which would remove an important subgroup of individuals with DS who are often underrepresented in research, a rank order scoring approach was used to characterize performances at the floor. Rank order conventions involved assigning values below the floor (e.g., "-2" and "-1") to differentiate participants who showed evidence of emerging skills on a given task (i.e. passing a practice item but not test items) versus no evidence of task engagement.

Working Memory/Short-Term Memory Measures

Garage Game (WM; Devine et al., 2019; Pinks et al., 2023). Children were presented with a toy set of cars and garages and were instructed to identify the location of hidden cars behind garage doors. Each garage held 3 different colored cars at the start of each trial and children were prompted to "Find a car!" In the practice trial, participants were able to see the cars they had found on the tabletop. In test trials, participants were prompted to wait to select a car between presentations of a distractor screen and the removal of the car from view after a correct search. Participants who passed practice were presented with up to three phases of the task. The first two test phases had 3 cars hidden, which were in the same garage as the practice trial. In the third phase, children were introduced to a second 3-car garage in addition to the garage from the previous practice and test trials. A repetitive search rate (RSR) was generated by calculating the number of incorrect searches participants made divided by the number of cars that the participant searched for (maximum of 12 cars for all 3 phases; Pinks et al., 2023). If a participant passed the practice trial, but they were unable to score on the test trials, they received a ranked score of "-1". If a participant did not pass the practice trial, they received a ranked score of "-2" for analyses. Scores were reverse coded after calculating the RSR to keep consistency in rank order scoring, such that higher scores indicated better performance (see Table 2 for ranges of performance). In addition to reverse scoring, 2 points were added to every score to have a range of performances from “0.00” to “3.00” for this measure. The Garage Game has shown moderate test-retest reliability (ICC = .60), and developmental sensitivity for children with DS 3 to 7 years (Pinks et al., 2023).

Childhood Modified Corsi Span Task (STM; Kessels et al., 2000; {withheld for review}, in press). As per {*withheld for review}* (in press), short-term memory performance was evaluated via the Corsi Span Task, wherein an examiner modeled tapping on an array of four large colorful button switches programmed to each produce a unique tone when pushed. The examiner then moved the tray toward the child to allow an immediate imitation of the sequence presented. Trials became increasingly complex, and when the trial sequence was greater than 2, children received scores of incorrect (“0”), correct buttons, incorrect order (“1”), or correct replication (“2”). Scores were summed based on total performance scores across all trials. Because of the complexity of the task and the wide CA range in the study, the extended scoring procedures were adopted so as not to exclude scores of zero correct, as this information is meaningful for establishing construct validity as well. For participants at the floor of the measure (no correct test items), scores of “-1” were assigned if the child imitated the sequence after one teaching opportunity and “-2” if the participant gave an incorrect response for both teaching opportunities. This task has demonstrated adequate feasibility and moderate retest reliability in this population (ICC = .62; {*withheld for review*}, in press). Notably, this task captures wide performance variability, a maximum score of 28 was observed in this investigation and a minimum score of -2.

Inhibitory Control Measures

Prohibition Sparkle Wand (Inhib1; Friedman et al., 2011; Walsh et al., 2023). Participants were presented with an enticing sparkle wand toy and were instructed to wait to contact the toy until the examiner said, “Go! You can play with it now!” after a 30-second waiting period. The sparkle wand was clear with glitter shapes and colored gel inside. This task has been previously evaluated for feasibility and developmental sensitivity in DS (Walsh et al., 2023). Participant latency to contact the toy was coded by research assistants naïve to study hypotheses. Coders established reliability for the coding scheme (IRR = .97; Noldus Information Technology, 2013). Latency scores from Observer 15XT were transformed into a categorical variable for analyses. The primary outcome measure was 4 categories of time to contact the toy (0 = Contact During Instructions, 1 = No waiting (0 – 5 seconds), 2 = Emerging Inhibition (6 – 29 seconds), 3 = Inhibited (30 seconds+), which allowed for the characterization of different levels of inhibitory control development. The feasibility of this measure has been demonstrated in previous investigations in DS, albeit with interpretational limitations at the youngest and oldest participant chronological ages (Walsh et al., 2023).

Prohibition Rainmaker (Inhib2; Friedman et al., 2011; Walsh et al., 2023). Similar to the sparkle wand prohibition task, children were presented with a clear wand with beads and moving pieces inside. As the rainmaker was tipped, beads moved around inside and made pieces spin and flip inside, with corresponding sounds of beads moving. Instructions and scoring procedures followed the process for Prohibition Sparkle Wand. All coders scored both tasks in the same project in Observer XT, however reliability was run separately for the tasks (Rainmaker IRR = .95; Noldus Information Technology, 2013). The sparkle wand and the rainmaker tasks were always administered separately within the order of task administration; previous evaluations have demonstrated that performances on the Sparkle Wand and Rainmaker tasks are moderately correlated (r = .50; Walsh et al., 2023).

Snack Delay (Inhib3; Carlson, 2005). The snack delay task involved the presentation of either a snack or toy of interest under a clear plastic cup. Participants were instructed to wait to retrieve either the snack or toy until the examiner rang the bell and said, "Go!". Four trials were presented with 5-, 10-, 20-, and 15- second wait times, respectively. Participants were not provided corrective feedback if they had an early retrieval of either the toy or snack. Participation in the task was coded by research assistants naïve to study objectives who observed the time to dysregulated behavior (e.g., tapping or contact with the cup, making contact with the bell, elopement). The mean latency to producing a dysregulated behavior was calculated, with longer durations indicating greater inhibition (see Table 2 for descriptive statistics). Coders maintained strong interrater reliability on the coding scheme (IRR = .85; Noldus Information Technology, 2013).

Cognitive Flexibility Measures

Adapted Reverse Categorization (Flex1; Carlson, 2005; Van Deusen et al., 2023). The Adapted Reverse Categorization task involved the sorting of red blocks labeled "ketchup" and yellow balls labeled "mustard". In the first pre-switch phase, participants sorted with a color-congruent rule wherein "ketchup" went into the "ketchup" bucket (i.e., red bucket with image of ketchup bottle), and the "mustard" was sorted into the "mustard" bucket (i.e., yellow bucket with image of mustard bottle). In the post-switch phase, participants sorted by an incongruent color rule. In this "silly game," children sorted "mustard" into the "ketchup" bucket and "ketchup" into the "mustard" bucket.

Participants could earn a maximum of 10 points per phase. In order to play the color-incongruent game, children had to pass a minimum of 6 trials on the color-congruent phase. The dimension of interest for this study was the number of correct trials in the color-incongruent phase. Children who passed fewer than three color-congruent trials were assigned a "-2" for analyses, and children scoring 4-5 correct in color-congruent were given a score of "-1". This increased the range of responses from -2 to 10. This task has shown test-retest reliability in a preliminary sample of 28 participants with DS (ICC = .81) and scalability across CAs 3 to 7 years and MAs of 1 to 3 years (Van Deusen et al., 2023).

Doll Stroop (Flex2; Bernier et al., 2010). Child participants were presented with both a small and large spoon and were asked to help feed a 'mommy' and a 'baby' cartoon pictures. Images were from a popular children's television show (*Peppa Pig*). In the first phase, children were instructed to use the small spoon to feed the baby pig, and the big spoon to feed the mommy pig. In the second phase, the "silly game", children were instructed to feed the mommy pig with the small spoon and the baby pig with the big spoon. Eight administrations were included in each phase. To play the silly game, six correct responses were required on the first phase. For inclusive scoring purposes, participants who did not meet this criterion were assigned scores for this assessment if they were successful on some of the pre-switch items. Children with three or fewer correct responses during the first phase were assigned a score of "-2", and children with 4-5 correct responses were assigned "-1".

Zoo Animal Switch (Flex3; Carlson, 2005). Children were instructed to sort big and small animals into big and small buckets. Participants were presented with safari animal pairs of big and small animals. During the first phase, children were instructed to sort the big animals into the big bucket (e.g., a large gray metal bucket), and the small animals into the small bucket (e.g., a small gray metal pail). Participants were presented with animal pairs where the small and big animals were both visible. One animal at a time was labeled and presented to the participant, (e.g., "This is a big animal, where does it go?") and the small animal stayed in the examiner's other hand until the first item was sorted. Big and small animals were presented in the opposite order for each animal pair (e.g., big then small for the first pair, small then big for the second pair), but the first animal was always presented from the examiner's left hand. This assessment procedure was intended to prevent sorting to one side meeting the minimum criteria to participate in the rule-change phase. If participants correctly sorted 6 or more animals, they moved on to the "silly game," where the big animals went into the small bucket and the small animals went into the big bucket. As with the other cognitive flexibility tasks, children scored on this measure even if they did not make it to the silly game. Trial 1 scores of 0 to 3 were assigned "-2", and scores of 4 and 5 were assigned "-1" in analyses. In each trial, participants were presented with five animal pairs, resulting in 10 opportunities to correctly sort in the silly game. Thirteen participants did not make it to the test trials due to refusal or noncompliance in teaching trials of this task.

Caregiver Report

Demographics and Medical History. Caregivers provided demographic (e.g., race, ethnicity) and biomedical information regarding their child (e.g., congenital heart defects, prematurity, and biomedical conditions).

Analytic Approach

The goal of the present study was to characterize the factor structure of an EF battery designed for young children with DS. The analytic approach involved the use of exploratory graph analysis (EGA; for complete description of network psychometrics see Epskamp et al., 2018) to examine the construct validity of the EF battery measures. Exploratory graph analysis (EGA) is a network approach used to identify the underlying factor structure of data (Golino et al., 2020). In recent simulations, Golino and Demetriou (2017) showed that EGA outperformed parallel analysis and the minimum average partial procedure for identifying the underlying dimensions of data. In this investigation, it was confirmed that parallel analysis identified the same structure. Their simulations demonstrated that EGA works well across a range of factor correlations and number of factors. EGA identifies the underlying factor structure by estimating an undirected network model using a partial correlation matrix. In the network model, nodes represent items and statistical relations among items are represented with edges. An advantage of EGA is the utilization of the least absolute shrinkage selection operator (LASSO) which suppresses small coefficients to zero over other network models (Tibshirani, 2018). Applying the LASSO penalty results in conditional independence among the factors and improves interpretability of the network model (Tibshirani, 2018). Additional benefits of EGA include immediate interpretation, eliminates researchers needing to decide the type of rotation in the factor structure, and reduces need for additional steps to verify factors compared to other dimensionality approaches (Golino et al., 2020).

The statistical software R was used for data cleaning, data visualization, and EGA analyses (Golino et al., 2023; R Core Team, 2024; van Lissa, 2023; Wickham et al., 2019). This EGA analysis used the Walktrap algorithm, Pearson correlation matrix, and missingness was handled through pairwise deletion. Unidimensional assessment was evaluated by Louvain, however with 2 identified clusters the 1-dimensional solution was not the best fit and thus the EGA approach was used (Jamison et al., 2021).

Results

Exploratory Graph Analysis

Table 2 reports mean performance, standard deviations, range, skewness, and kurtosis for each EF task. EGA results are presented in Figure 1, and the partial correlation matrix is presented in Table 3. In the partial correlation matrix, values of 0 indicate associations that were suppressed to 0 by the LASSO penalty. The non-suppressed correlations ranged from small to large. The EGA identified two latent factors with positive associations between the factors. The lines indicate positive associations between the task outcomes, and line thickness is indicative of the strength of the association. The two factors demonstrated clear distinguishability. One factor included all three inhibitory control assessment performances and was named the “Inhibition” (Inhib) factor. The other factor included measures of cognitive flexibility, working memory, and short-term memory, and was named the “Working Memory/Flexibility” (WM/Flex) factor.

Discussion

The present study aimed to evaluate the factor structure of a newly compiled EF battery designed for use with young children with Down syndrome. Currently available EF assessments often require the use of additional developmental skills like language and motor planning, thus presenting interpretational confounds and compromising the validity of the assessments. The EXcEEDS battery was designed to facilitate valid interpretations and uses of EF assessment in DS as part of the larger effort to evaluate outcome measures for use in future treatment studies in this population. The EGA analyses presented in this study demonstrated that EXcEEDS battery measures load onto two latent factors in ways that are meaningfully interpretable. Notably, all three inhibitory control measures loaded onto a single latent factor, with no other

EF tasks loading onto this dimension. A second latent factor included the remaining measures designed to assess flexibility and foundations of working memory.

EF is a critical target for early intervention in DS, but novel treatments cannot be evaluated for efficacy without valid, reliable, and developmentally sensitive outcome measures. Several EXcEEDS measures included in the present study have previously demonstrated psychometric soundness along the dimensions of developmental sensitivity, scalability, minimal floor effects, minimal practice effects, and preliminary evidence of test-retest reliability ({withheld for review}, in press; Pinks et al., 2023; Van Deusen et al., 2023; Walsh et al., 2023). The present study contributes to this validation effort by providing additional evidence of the underlying construct validity for the measures included in the battery.

An important element of the present study was the consideration of EF assessment that sought to reduce confounding variables to isolate EF performances. To achieve this, EXcEEDS battery measures were developed by adapting currently available laboratory EF assessments designed for young children in the general population. Although there is debate regarding the underlying factor structure of EF, most models include the constructs examined in the present study (Miyake et al., 2000; Müller & Kerns, 2015). Adaptations to widely used measures of early childhood working memory, inhibition, and flexibility involved minimizing syndrome-related interpretational and measurement confounds, including eliminating expressive language response requirements, reducing receptive language demands, and minimizing motor planning complexity. The present study builds on previous evaluations of individual tasks to establish the construct validity across the collective battery and the convergence of task performances along dimensions of interest.

Early EF architecture in DS. Although the primary aim of this investigation was to examine the psychometrics of a novel EF battery for young children with DS, results may also inform the ongoing scientific discussion regarding the unitary versus dissociable nature of EF and the dissociability of EF components during early development (Wiebe et al., 2011). The present study introduces novel information into this broader discussion, with evidence that inhibitory aspects of EF in young children with DS appear to be dissociable from the constructs of working memory and flexibility. This finding provides evidence for some degree of early EF dissociability, but also evidence of a degree of unitary executive control in the alignment between working memory and flexibility task performances. This novel finding warrants replication, but if reproducible, it may contribute to the understanding of early EF architecture in clinical populations.

In the context of this novel contribution to the study of EF architecture, it is also important to consider alternative interpretations for the latent factor structure observed. It may be the case that the inhibitory control tasks differed from the working memory and flexibility tasks in other systematic ways that are not reflective of underlying EF development. For example, the working memory and flexibility tasks each required the child participants to produce some type of motoric response (e.g., gesturing toward a toy, pressing a button), and the three inhibitory control measures involved a child *not* producing any such responses. Thus, it could be plausible to explain the underlying factor structure along the dimension of the child's degree of initiation or responsivity to external stimuli. However, the action plans required for the working memory and flexibility tasks did not involve the production of just any type of response. These tasks required increasingly more complex responses (e.g., longer spans, correct post-switch sorting) for increased scores, and producing an activated and incorrect response would not earn higher scores on this category of tasks. The fact that correct responses across all of these tasks tended to be associated with one another suggests that some underlying cognitive regulatory process undergirded performances and not individual differences in initiation of any type.

Next Steps. Further evaluation of the EXcEEDS battery is warranted along the dimensions of clinical meaningfulness, developmental sensitivity, and utility for other clinical populations. Preliminary evidence for the clinical meaningfulness of the EXcEEDS battery has been demonstrated in a recent study that established a predictive association between performance on a selection of EXcEEDS measures and adaptive behavior outcomes one year later (*citation withheld for review*). Additional research, however, is needed to quantify the association between EF performances and other neurodevelopmental outcomes that co-occur in DS, and that also are known to impact EF, such as autism and attention deficit hyperactivity disorder. In particular, clinical meaningfulness could be demonstrated if EXcEEDS battery performances differed for children with DS who do and those who do not have co-occurring neurodevelopmental conditions. If this were to be observed, EXcEEDS performances could even be a potentially useful tool to signal the need for additional neuropsychological evaluations. Future efforts to evaluate clinical meaningfulness may also inform the convergent and discriminant validity of the EXcEEDS battery. Additional evaluations of the association between EXcEEDS performances and performances on other EF direct assessments and proxy report measures would be beneficial for evaluating these validity dimensions as well.

Future evaluations should also quantify the degree to which EXcEEDS measures demonstrate sensitivity to change. Cross-sectional associations between CA, MA, and EXcEEDS task performances have been reported elsewhere (Pinks et al., 2023; {withheld for review}; Van Deusen et al., 2023; Walsh et al., 2023). However, previous cross-sectional findings should not be conflated with longitudinal changes. Therefore, another important next step would be to evaluate performances throughout the course of a year or longer interval to determine whether within-individual changes are captured.

Although this investigation focused on understanding the utility of the EXcEEDS Battery in children with DS, future measurement evaluation efforts should also assess the use of this battery with participants in this developmental range without DS, including those with other neurogenetic conditions and those without disabilities. An important starting point for the development of this battery was to adapt it to the phenotype associated with DS. The current lack of validated measures for this population has made intervention and treatment challenging, as valid measures are needed to determine whether treatments are effective or not (Baumer et al., 2022; Esbensen et al., 2017). With measurement validation occurring in this range of early childhood CAs and MAs in DS, replication and extension to additional populations at comparable developmental levels is needed to validate the broader use of the EXcEEDS Battery.

Limitations

Although the sample size in the present study is substantial for neurogenetic syndrome research, the sample is relatively smaller than those reported in the broader psychometric evaluation literature. It is notable, however, that a smaller sample size did not preclude the identification of a strongly fitting model for the psychometric evaluation of the EXcEEDS battery. Another important consideration is the homogeneity of the sample in terms of race, ethnicity, and income distribution. Future work should aim to include more diverse participants. Recent work from an NIH Outreach Working Group has highlighted methods including the use of a community based participatory research approach, increasing the accessibility of research by helping the place of research be closer to families, and increasing the dissemination of research findings in lay terms to directly share current work with participating families (Fidler et al., 2022).

The present investigation aimed to include all participants in analyses, which motivated the use of an adapted rank-order scoring procedure. This novel approach potentially addresses the issue of heterogeneity of performance among children with DS and prevents the exclusion of children with more pronounced delays from group-level findings. However, this is a newly adapted approach to preventing such participant exclusion, and therefore, should be interpreted with some measure of caution. Although the addition of rank-ordered scoring increased the range of several tasks, there are some limitations of those outcome measures with a narrow range of response options (i.e., inhibition and working memory). Future work should evaluate the reliability of EXcEEDS composite scores and continue to monitor how the range of possible responses influences this critical dimension.

Another important consideration for this study is related to the measurement of working memory. The early measurement of working memory is often interrelated with early short-term memory foundations. While short-term memory entails the temporary storage of visual-spatial or verbal information, working memory includes an active manipulation or updating component (Conners et al., 2011). However, working memory and short-term memory appear to be less dissociable in early development, and "simple working memory" tasks utilized in infant research often resemble measures of short-term memory (Bell et al., 2022). As such, the EXcEEDS battery was designed to include both a measure of working memory that involved updating, as well as a span task that involved reproducing sequences without manipulation. This approach was justified by the need to include a measure of memory capacity beyond updating, but in the context of overall developmental level in the study CA range, wherein span tasks that require manipulation (e.g., backward span) are too developmentally complex. It is also notable that the EXcEEDS battery model included only two working memory-related activities, which precluded the identification of an additional separable subcomponent because of the inclusion of fewer than three indicator variables. In future work, the evaluation of additional measures of working memory and its foundations will be a priority to determine the separability of working memory and flexibility task performances, which loaded on to the same latent factor in the present study.

Despite these limitations, the present study contributes new information regarding psychometrically sound approaches to the measurement of EF in young children with DS. Results suggest that adapting EF tasks to reduce receptive language, expressive language, and motor demands is a promising strategy for measuring EF in this population. Future work should continue to expand the evidence base for the utility of the EXcEEDS battery by establishing age-related norms for children with DS and by identifying the longitudinal predictive utility of EF task performances over time.

References

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: The predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *25*(3), 157–173.

Baumer, N. T., Becker, M. L., Capone, G. T., Egan, K., Fortea, J., Handen, B. L., Head, E., Hendrix, J. E., Litovsky, R. Y., Strydom, A., Tapia, I. E., & Rafii, M. S. (2022). Conducting clinical trials in persons with Down syndrome: summary from the NIH INCLUDE Down syndrome clinical trials readiness working group. *Journal of Neurodevelopmental Disorders*, *14*(1), 22.

Bell, M. A., Phillips, J. J., & Bruce, M. D. (2022). Infant and toddler working memory. In *The development of memory in infancy and childhood* (pp. 87–110). Psychology Press.

Bernier, A., Carlson, S. M., & Whipple, N. (2010). From External Regulation to Self-Regulation: Early Parenting Precursors of Young Children's Executive Functioning. *Child Development*, *81*(1), 326–339.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616.

Channell, M. M., Mattie, L. J., Hamilton, D. R., Capone, G. T., Mahone, E. M., Sherman, S. L., Rosser, T. C., Reeves, R. H., Kalb, L. G., & Down Syndrome Cognition Project. (2021). Capturing cognitive and behavioral variability among individuals with Down syndrome: A latent profile analysis. *Journal of Neurodevelopmental Disorders*, *13*(1), 16.

Conners, F. A., Moore, M. S., Loveall, S. J., & Merrill, E. C. (2011). Memory profiles of Down, Williams, and fragile X syndromes: Implications for reading development. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *32*(5), 405–417.

Devine, R. T., Ribner, A., & Hughes, C. (2019). Measuring and predicting individual differences in executive functions at 14 months: A longitudinal study. *Child Development*, *90*(5), e618–e636.

Engelhardt, L. E., Harden, K. P., Tucker-Drob, E. M., & Church, J. A. (2019). The neural architecture of executive functions is established by middle childhood. *NeuroImage*, *185*, 479–489.

Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2018). Network Psychometrics. In *The Wiley Handbook of Psychometric Testing* (pp. 953–986). John Wiley & Sons, Ltd.

Esbensen, A. J., Hooper, S. R., Fidler, D., Hartley, S. L., Edgin, J., D'Ardhuy, X. L., Capone, G., Conners, F. A., Mervis, C. B., Abbeduto, L., Rafii, M., Krinsky-Mchale, S. J., Urv, T., Dykens, E., Esbenson, A., Hartlay, S., Keller, S., & Weir, S. (2017). Outcome measures for clinical trials in Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *122*(3), 247–281.

Fidler, D. J., Riggs, N., Esbensen, A. J., Jackson-Cook, C., Rosser, T., & Cohen, A. (2022). Outreach and engagement efforts in research on Down syndrome: An NIH INCLUDE Working Group Consensus Statement. *International Review of Research in Developmental Disabilities*, *63*, 247–267.

Fidler, D. J., Van Deusen, K., Prince, M. A., Schworer, E. K., Lee, N. R., Edgin, J. O., Patel, L. R., & Daunhauer, L. A. (2023). Longitudinal predictors of neurodevelopmental outcomes in children with Down syndrome. *Developmental Neuropsychology*, 1–19.

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*(5), 1410–1430.

Golino, H., Christensen, A. P., & Moulder, R. (2023). EGAnet: Exploratory graph analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics. *R Package Version 0. 9*, *5*.

Golino, H., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence*, *62*, 54–70.

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292–320.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Biomedical Informatics Insights*, *42*(2), 377–381.

Hessl, D., Sansone, S. M., Berry-Kravis, E., Riley, K., Widaman, K. F., Abbeduto, L., Schneider, A., Coleman, J., Oaklander, D., Rhodes, K. C., & Gershon, R. C. (2016). The NIH Toolbox Cognitive Battery for intellectual disabilities: three preliminary studies and future directions. *Journal of Neurodevelopmental Disorders*, *8*(1), 35.

Jamison, L., Christensen, A. P., & Golino, H. (2021). *Optimizing Walktrap's community detection in networks using the total entropy fit index*. https://osf.io/preprints/psyarxiv/9pj2m/

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258.

Lanfranchi, S., Jerman, O., Dal Pont, E., Alberti, A., & Vianello, R. (2010). Executive function in adolescents with Down Syndrome. *Journal of Intellectual Disability Research: JIDR*, *54*(4), 308–319.

Loveall, S. J., Conners, F. A., & Tungate, A. S. (2017). A cross-sectional analysis of executive function in Down syndrome from 2 to 35 years. *Journal of Intellectual Disability Research*. https://onlinelibrary.wiley.com/doi/abs/10.1111/jir.12396

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Müller, U., & Kerns, K. (2015). The development of executive function. In L. S. Liben (Ed.), *Handbook of child psychology and developmental science: Cognitive processes, Vol* (Vol. 2, pp. 571–623). John Wiley & Sons, Inc., xxv.

Noldus Information Technology. (2013). *Noldus Observer XT* (Version 15) [Computer software]. http://www.noldus.com/human- behavior-research/products/the-observer-xt

Ozonoff, S., & Jensen, J. (1999). Brief report: Specific executive function profiles in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders*, *29*(2), 171–177.

Pinks, M. E., Van Deusen, K., Prince, M. A., Esbensen, A. J., Thurman, A. J., Patel, L. R., Abbeduto, L., Walsh, M. M., Daunhauer, L. A., Feigles, R. T., Nguyen, V., & Fidler, D. J. (2023). Psychometric evaluation of a working memory assessment measure in young children with Down syndrome. *Research in Developmental Disabilities*, *139*, 104564.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schworer, E. K., Altaye, M., Fidler, D. J., Beebe, D. W., Wiley, S., Hoffman, E. K., & Esbensen, A. J. (2023). Evaluating Processing Speed and Reaction Time Outcome Measures in Children and Adolescents with Down Syndrome. *International Journal of Environmental Research and Public Health*, *20*(6). https://doi.org/10.3390/ijerph20065202

Tibshirani, R. (2018). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *58*(1), 267–288.

Tomaszewski, B., Fidler, D. J., Talapatra, D., & Riley, K. (2018). Adaptive behaviour, executive function and employment in adults with Down syndrome. *Journal of Intellectual Disability Research: JIDR*, *62*(1), 41–52.

Tungate, A. S., & Conners, F. A. (2021). Executive function in Down syndrome: A meta-analysis. *Research in Developmental Disabilities*, *108*, 103802.

Van Deusen, K., Prince, M. A., Thurman, A. J., Esbensen, A. J., Patel, L. R., Abbeduto, L., Walsh, M. M., Daunhauer, L. A., Tempero Feigles, R., & Fidler, D. J. (2023). Evaluating an adapted reverse categorization task to assess cognitive flexibility in young children with Down syndrome. *Journal of Intellectual Disability Research: JIDR*. https://doi.org/10.1111/jir.13040

van Lissa, C. J. (2023). _*tidySEM: Tidy structural equation modeling* (0.2.4) [R package]. ttps://CRAN.R-project.org/package=tidySEM>.

Walsh, M. M., Van Deusen, K., Prince, M. A., Esbensen, A. J., Thurman, A. J., Pinks, M. E., Patel, L. R., Feigles, R. T., Abbeduto, L., Daunhauer, L. A., & Fidler, D. J. (2023). Preliminary psychometric properties of an inhibition task in young children with Down syndrome. *Journal of Intellectual Disabilities: JOID*, 17446295231218776.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436–452.

Will, E. A., Fidler, D. J., Daunhauer, L., & Gerlach-McDonald, B. (2017). Executive function and academic achievement in primary - grade students with Down syndrome. *Journal of Intellectual Disability Research: JIDR*, *61*(2), 181–195.

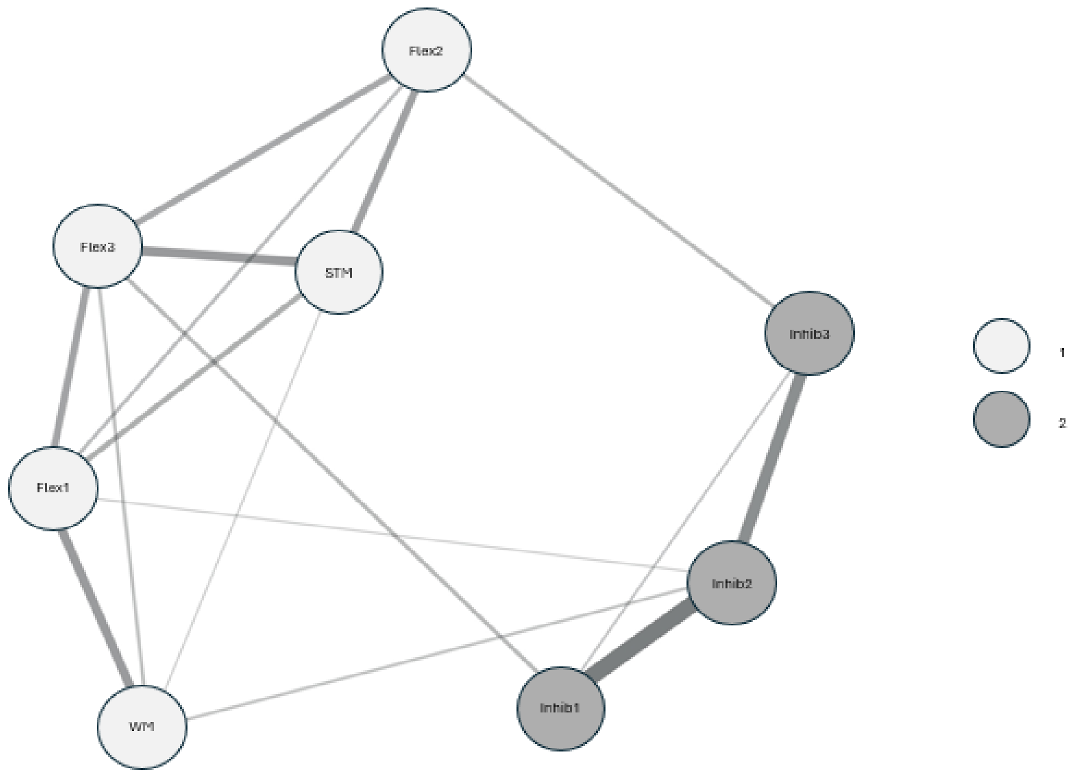
Willoughby, M., & Hudson, K. (2021). Current issues in the conceptualization and measurement of executive function skills. *Executive Functions and Writing*, 17.

Yang, Y., Conners, F. A., & Merrill, E. C. (2014). Visuo-spatial ability in individuals with Down syndrome: Is it really a strength? *Research in Developmental Disabilities*, *35*(7), 1473–1500.

Figure

Click here to access/download;Figure;Figure 1. Exploratory Graph Analysis_updated.docx

Figure 1. Exploratory Graph Analysis



Note: Modified Corsi Total Score = STM; Garage Game Repetitive Search Rage = WM; Sparkle Wand Latency to Retrieve = Inhib1; Rain Maker Latency to Retrieve = Inhib2; Snack Delay Average Latency to Dysregulated Behavior = Inhib3; Adapted Reverse Categorization Post Switch Correct = Flex1; Doll Stroop Post Switch Correct = Flex2; Zoo Animal Switch Post Switch Correct = Flex3; Dark gray nodes represent the latent construct of "Inhibition"; Light gray ed nodes represent the latent construct of "Flexibility". Thicker lines indicate a stronger association. Thinner lines represent a weaker association. Missing lines represent associations that were suppressed to 0 by the LASSO penalty.

 Click here to access/download;Table;Table 1. Demographic Information.docx

Table 1. Demographic Information

Child variable	**% (n)**
% Male	52.4 (65)
Child Chronological Age (years)	5.26 (1.59)
Child Mental Age (years; *n* = 8 missing)	2.37 (0.91)
Race (*n* = 8 missing)	
American Indian or Alaska Native	0.9 (1)
Asian-American	3.4 (4)
Black/African-American	1.7 (2)
White	86.2 (100)
Other	7.8 (9)
Ethnicity (*n* = 15 missing)	
Hispanic	16.5 (18)
Not Hispanic	83.5 (91)
DS Type (*n* = 7 missing)	
Trisomy 21	88.0 (103)
Mosaicism	1.7 (2)
Translocation	5.1 (6)
Not Sure	5.1 (6)
Premature Birth (% yes; *n* = 7 missing)	26.5 (31)
Congenital Heart Defects (% yes; *n* = 7 missing)	70.9 (83)
Caregiver variable	
Primary Caregiver Age (Mean/SD; *n* = 7 missing)	40.79 (6.48)
% Primary Caregiver Education at least 1 year of college/tech training (n; *n*= 9 missing)	93.9 (108)
% Annual Household Income (n; *n* = 7 missing)	
Below $50,000	6.8 (8)
$50,000-100,000	29.9 (35)
Above $100,000	60.7 (71)
Did not wish to provide	2.6 (3)

Click here to access/download;Table;Table 2. Executive Function Task Performances.docx

Table 2. Executive Function Task Performance Descriptive Statistics.

Task	N	Mean (SD)	Range	Skewness	Kurtosis	Percent of Participants at Floor (n)		Percent of Participants at Ceiling (n)
						-2	-1	
STM	119	4.43 (7.49)	-2 - 28	1.24	0.80	31.93 (38)	5.88 (7)	0
WM	122	2.17 (1.03)	0 - 3	-1.29	0.12	14.75 (18)	4.92 (6)	17.21 (21)
Inhib1	110	1.34 (0.99)	0 - 3	0.54	-0.79	17.27 (19)		20 (22)
Inhib2	109	1.37 (1.04)	0 - 3	0.45	-1.03	19.26 (21)		22.94 (25)
Inhib3	113	7.29 (6.22)	0.03 - 21.12	0.51	-1.17	0		0
Flex1	114	3.61 (4.87)	-2 - 10	0.31	-1.69	14.04 (16)	8.77 (10)	28.07 (32)
Flex2	112	0.92 (3.56)	-2 - 8	1.00	-0.62	33.93 (38)	18.75 (21)	8.93 (10)
Flex3	111	2.40 (4.300)	-2 - 10	0.73	-0.97	12.61 (14)	29.73 (33)	17.12 (19)

Click here to access/download;Table;Table 3. EGA partial correlation matrix with LASSO.docx

Table 3. EGA partial correlation matrix with LASSO penalty applied.

	STM	WM	Inhib1	Inhib2	Inhib3	Flex1	Flex2	Flex3
STM	-							
WM	0.06	-						
Inhib1	0.00	0.00	-					
Inhib2	0.00	0.10	0.52	-				
Inhib3	0.00	0.00	0.09	0.37	-			
Flex1	0.17	0.28	0.00	0.06	0.00	-		
Flex2	0.24	0.00	0.00	0.00	0.13	0.12	-	
Flex3	0.29	0.11	0.13	0.00	0.00	0.23	0.20	-

Note: Modified Corsi Total Score = STM; Garage Game Repetitive Search Rage = WM; Sparkle Wand Latency to Retrieve = Inhib1; Rain Maker Latency to Retrieve = Inhib2; Snack Delay Average Latency to Dysregulated Behavior = Inhib3; Adapted Reverse Categorization Post Switch Correct = Flex1; Doll Stroop Post Switch Correct = Flex2; Zoo Animal Switch Post Switch Correct = Flex3. Values indicate the partial correlations used in the exploratory graph analysis (EGA). Values fixed to zero were fixed by the LASSO penalty. Effect size rules of thumb for correlations: .1 = small, .3 = medium, .5 = large.