Mapping neurodevelopmental diversity in executive function

Silvana Mareva[1,2], The CALM Team[1], & Joni Holmes[1,3*]

[1] MRC Cognition & Brain Sciences Unit, University of Cambridge, Cambridge, UK

[2]Faculty of Life Sciences, University of Exeter, UK

[3]School of Psychology, University of East Anglia, UK

*Corresponding author: Joni Holmes, School of Psychology, University of East Anglia, Norwich, NR4 7TJ, UK. E-mail: joni.holmes@uea.ac.uk. Tel: +44 1603 456161

Abstract

Executive functions can be conceptualised as either a set of higher-order cognitive skills that enable us to engage in flexible thinking and regulate our thoughts and behaviours, or as the ability to integrate knowledge, beliefs, and values when applying cognitive control in everyday situations. These two perspectives map onto the ways in which executive function is measured in childhood – using either structured laboratory tasks or ratings of everyday behaviours. Differences in executive functioning are associated with neurodevelopmental differences, but evidence for associations between specific profiles of executive function and specific neurodevelopmental conditions is mixed. In this study, we adopt a data-driven approach to identify common profiles of executive function in a transdiagnostic sample of 566 neurodivergent children, using both performance and rating-based measures of executive function. Three profiles of executive function were identified: one had consistent difficulties across both types of assessments, while the other two had inconsistent profiles of predominantly rating- or predominantly task-based difficulties. Children with these different profiles had differences in academic achievement and mental health outcomes and could further be differentiated from a comparison group of neurotypical children on both shared and profile-unique patterns of neural white matter organisation. Importantly, children's executive function profiles were not directly related to diagnostic categories or to dimensions of neurodiversity associated with specific diagnoses (e.g., hyperactivity, inattention, social communication). These findings support the idea that there are separate domains of executive function and that the two types of assessment tapping these functions are dissociable and provide non-redundant information related to neurodevelopmental differences. These findings advance our understanding of executive function profiles in neurodivergent populations and their relationship to behavioural outcomes and neural variation.

Key words: executive function, neurodiversity, neurodevelopment

Introduction

The transdiagnostic revolution in neurodevelopmental conditions emphasises a move away from diagnostic systems towards identifying broad dimensions that characterise neurodiversity (Astle et al., 2021). One such dimension, which captures characteristics associated with multiple neurodevelopmental outcomes (e.g., Bloemen et al., 2018) and can be a barrier to learning (Holmes et al., 2020) is executive function. The term executive function is used widely to describe the processes supporting the volitional control of cognition and behaviour. In this study, we adopt a transdiagnostic data-driven approach to delineate profiles of executive function among a large neurodevelopmentally neurodivergent sample using measures designed to capture two conceptually distinct aspects of executive function: performance-based tasks designed to tap higher-order cognitive skills (Miyake et al., 2000), and behaviour ratings suited to capturing the application of cognitive control in everyday contexts (Doebel, 2020). We then explore whether and how these data-driven groups are associated with differences in neurodevelopmental diagnostic categories and dimensions, academic and mental health outcomes, and neural white matter organisation.

Executive function can be conceptualised in two ways that map broadly on to the ways in which they are measured (Malanchini et al., 2018; Toplak et al., 2013). The first views executive function as a set of higher-order cognitive skills that enable flexible thinking (Miyake et al., 2000). These abilities include shifting/switching, inhibition, working memory, planning, and sustained attention (Friedman & Miyake, 2017; Miyake et al., 2000; St Clair-Thompson & Gathercole, 2006a). They are typically indexed by accuracy or response times on performance-based tasks. These include memory span tasks that index working memory performance (e.g., Automated Working Memory Assessment, Alloway, 2007), tasks that require goal-oriented planning (e.g., Tower of London, Shallice, 1982), semantic incongruence or cognitive flexibility tasks that assess shifting/switching (e.g., Number-Letter sequencing, Delis et al., 2001; Stroop, Stroop, 1935), and tasks that require focused attention in the face of distraction over a lengthy periods that measure sustained attention and response inhibition (e.g., Continuous Performance Test , Conners, 1992). There are many other task-based measures of executive function, and multiple variants of the examples provided here (see Strauss et al., 2006 for a

comprehensive list of task-based measures of executive function), but what they have in common is that they are administered individually under controlled experimental conditions requiring effortful cognitive control. In this way, therefore they can be viewed to reflect optimal rather than typical performance.

The second view of executive function integrates knowledge, beliefs, and values alongside cognitive, motor, and perceptual control, and makes explicit that these factors come together in the service of goals in specific settings (Doebel, 2020). It refers to the individual use of executive skills in particular situations, rather a general "capacity process" such as inhibitory control that is applied across all situations. To illustrate the distinction, Doebel (2020) uses the example of a child inhibiting the impulse to hit another child who has taken their toy. In this situation, the child brings to bear knowledge and beliefs about hitting another child (e.g., socially acceptable alternatives to hitting, beliefs about being scolded for engaging in hitting behaviour) when exerting inhibitory control. In this way, executive functions are viewed as inseparable from the task goals, context, and mental content that comes into play in specific situations. Understanding executive function as a concept that is tied to the activities a person is completing in their everyday activities mirrors the second way in which executive function is measured – using rating scales to index the ability to coordinate multiple processes in everyday problem-solving situations that draw on knowledge, beliefs, and values. Examples of these rating scales include the Behaviour Rating Index of Executive Function (BRIEF; Gioia et al., 2000) and the Working Memory Rating Scale (Alloway et al., 2009). These, and similar questionnaires, require observers to rate the frequency of everyday executive function behaviours in different contexts. They were originally developed to provide an ecologically valid measure of executive function that captures real-world functioning. Questionnaire responses typically reflect behavioural observations over extended periods of time, and may therefore reflect more trait-like or stable processes and provide a more accurate assessment of typical performance (Malanchini et al., 2018).

Two lines of evidence suggest that executive function in these two domains – cognitive processes or skills involved in goal-oriented behaviour in everyday situations – is dissociable. First, there is little convergent validity between the measurement types tapping each domain (Dang et al.,

2020). If performance-based and rating-based measures of executive function were assessing the same underlying construct, they should be highly correlated. Yet concurrent and predictive validity studies consistently indicate that they are, at best, weakly correlated (e.g., Gerst et al., 2017; Nin et al., 2022; Soto et al., 2020; Tamm & Peugh, 2019; Toplak et al., 2013), and that these relationships remain small when using latent variable approaches (Snyder et al., 2021) and/or when controlling for mono-method bias (e.g., that performance-based measures predict other performance-based measures, and rating-based measures predict other rating-based measures Soto et al., 2020). In a review, Toplak and colleagues (2013) concluded that the two types of measurement capture different information, with rating-based measures capturing goal-oriented successes in typical everyday settings and performance-based measures capturing processing-efficiency under optimal conditions.

Second, they make independent contributions to clinical, academic, and life outcomes. Difficulties with executive function are associated with multiple neurodevelopmental conditions, including attention deficit hyperactivity disorder (ADHD), specific learning difficulties, and autism (e.g., Benallie et al., 2021; Holmes et al., 2014; Loe & Feldman, 2007; McClain et al., 2022; Willcutt et al., 2005), they can be a barrier to learning (Holmes et al., 2020; Peng & Fuchs, 2016; Soto et al., 2021; Yeniad et al., 2013), and are also linked to poor mental health outcomes (e.g., Bloemen et al., 2018). Substantial evidence suggests that the different measurement methods corresponding to each domain of executive function make unique and separable contributions to these outcomes (Gerst et al., 2017; Soto et al., 2020). In terms of academic achievement, both rating- and performance-based measures predict outcomes, but the relationships are typically stronger and more consistent for performance-based measures (Gerst et al., 2017; Malanchini et al., 2018; Soto et al., 2020). Similarly, for mental health outcomes, rating- and task-based measures of executive function explain independent variance in both internalising and externalising difficulties (Eisenberg et al., 2019; Ellingson et al., 2019; Friedman et al., 2020; Friedman & Gustavson, 2022), with the stronger relationships typically reported between rating-based measures and mental health (Friedman et al., 2020).

The correspondence between ratings of executive function behaviours and performance on cognitive tests is also low in neurodivergent groups. For example, in individuals with ADHD, the

executive function profiles captured on performance-based and rating-based measures have little overlap (Biederman et al., 2008). Moreover, both types of assessments have their own distinct contribution to predicting occupational outcomes in this population (Barkley & Murphy, 2010, 2011). In a transdiagnostic study, Williams and colleagues (2022) found that a substantial proportion of a large sample of neurodivergent children had what they termed as an inconsistent executive function, defined as difficulties reported on rating-based assessments of executive function that were not evident on performance-based tasks.

Current study

The aim of the current study was to adopt a transdiagnostic approach to delineate profiles of executive function among a large neurodevelopmentally neurodivergent sample. Previous studies exploring executive function in neurodivergent children have typically either contrasted executive function performance between groups (e.g., those with a diagnosis compared those with a different, or no, diagnosis) across a range of performance-based measures (e.g., Geurts et al., 2004; Holmes et al., 2014), explored how well executive function rating scales and performance tests predict outcomes separately (e.g., Gerst et al., 2017; Soto et al., 2020), or tested relationships between different types of executive function measurement (e.g., Nin et al., 2022; Tamm & Peugh, 2019). Here, we adopted a different approach to explore individual profiles of executive function across both performance- and rating-based measures that were designed to tap into the cognitive process-based and situational goal-oriented conceptualisations of executive function. We used a self-organising map algorithm to map the multidimensional space of a broad range of performance and rating-based assessments of executive function, representing how children group together based on their executive function profiles. We then used data-driven clustering to delineate subgroups of children presenting with different executive function profiles and explored how these related to the two dissociable conceptualisations of executive function. We also investigated whether the groups differed in terms of neurodevelopmental diagnostic categories (ADHD, autism, etc.) and dimensions that capture neurodiversity (hyperactivity, inattention, social communication). Finally, we explored how these groups related to learning, mental health, and neural white matter organisation. We focussed on white matter organisation because there is a

substantial literature supporting its links to both executive function (Bathelt et al., 2018; Baum et al., 2017) and neurodevelopmental conditions and dimensions (Ameis et al., 2016; Beare et al., 2017). We applied this approach to data from a transdiagnostic sample that included a range of children who had been identified as having additional needs by health and education practitioners, irrespective of their diagnostic status, consistent with the notion that diagnostic-based studies fail to capture broad populations of neurodivergent children (Astle et al., 2021). The study was fully exploratory, and we did not formulate a hypothesis about the number of groups or the phenotypic and neural features that would differentiate them. Instead, we used a data-driven approach to address the following questions: 1) can we identify subgroups of children with different profiles of executive function; 2) do groups differ in terms the two dissociable conceptualisations of executive function; and 3) how do these groups relate to neurodevelopmental diagnostic characteristics, academic and mental health functioning, and neural white matter organisation?

Method

Participants

Participants were drawn from the Centre for Attention, Learning, and Memory (CALM) cohort. Recruitment details and testing procedures are described in the study protocol (Holmes et al., 2019). Broadly, children aged 5-18 years were referred to the study by health and educational professionals for difficulties with attention, learning, and/or memory regardless of diagnostic status. Participants aged 8 years and above were included in the current study. Younger children were excluded because some of the executive function tasks were not standardised for children under 8 years and were therefore not administered to them. The sample for this study included 566 children (45% diagnosed) with an average age of M = 10.55 years, SD = 2.02 (N_{boys} = 380, N_{girls} = 186). The most common diagnoses were ADHD (N=156), Autism (N=46), and dyslexia (N= 47). Ethical approval was obtained by the National Health Service (REC: 13/EE/0157). Parents/caregivers gave written consent and children gave verbal assent to participate.

Measures

Children completed a range of assessments of learning and cognition and a questionnaire asking about their mental health. Parents/caregivers completed questionnaires about the child's behaviour, communication skills, and mental health. All assessments were administered by a trained researcher at the CALM clinic, following standardised administration procedures documented in the test manuals. Task details and administration procedures are available in the study protocol and summarised briefly below.

Executive Function

Task-based measures - Tasks tapping executive functioning were chosen from the broader neuropsychological battery used in the CALM study (Holmes et al., 2019). These included four tasks from the Automated Working Memory Assessment (AWMA; Alloway, 2007), tapping verbal and visuospatial short-term/working memory: forward digit recall, backward digit recall, dot matrix, and Mr X. Two tasks from the Delis-Kaplan Executive Function System test battery (D-KEFS, Delis et al., 2001) were included: the Tower task, commonly used to assess planning skills, and the Trail-making number-letter sequencing task, a Stroop-like task indexing shifting. Two tasks from the Test of Everyday Attention for Children 2 (TEA-Ch2, Manly et al., 2016) were also included: the switching task, Reds, Blues, Bags and Shoes (RBBS), and the Vigil task tapping sustained attention.

Rating-based measures - Parents/caregivers completed the BRIEF (Gioia et al., 2000), an 80-item rating scale covering eight domains: Inhibition, Shifting, Emotional control, Initiation, Working memory, Planning, Organisation, and Monitoring.

Learning

The Word Reading and Numerical Operations subtests of the Wechsler Individual Achievement Test II (WIAT-II; Wechsler, 2005) were used to assess children's reading and mathematical skills.

Mental health

The total anxiety and depression score from the Revised Child Anxiety and Depression Scale (RCADS, Chorpita et al., 2005) was used to index child-reported internalising difficulties. This questionnaire was

introduced into the CALM protocol mid-way through the study, meaning data were only available for 269 participants.

Diagnostic categories and dimensions of neurodiversity

Neurodevelopmental diagnoses were reported by the referring professional and confirmed by parents/ caregivers at the time of referral to the CALM study. Ratings on dimensions of neurodiversity that are core features of neurodevelopmental conditions that were included in the CALM protocol were included in the current study. These were ratings provided by the child's parent / caregiver using the following scales: inattention and hyperactivity/impulsivity (dimensions of ADHD measured by the Conners-3 Parent Short Form Rating Scale (Conners, 2008); communication skills (dimension of developmental language disorder measured by the Child Communication Checklist-2, CCC-2, Bishop, 2003), and social communication and interests (core dimensions of autism, measured by the CCC-2, Bishop, 2003).

MRI data acquisition and pre-processing

All participants were invited to participate in an optional magnetic resonance imaging (MRI) session. T1-weighted volume scans were acquired using a whole-brain coverage 3D Magnetization Prepared Rapid Acquisition Gradient Echo (MP RAGE) sequence acquired using 1 mm isometric image resolution. Diffusion scans were obtained using echo-planar diffusion-weighted images with an isotropic set of 68 noncollinear directions. Whole brain coverage was based on 60 contiguous axial slices and isometric image resolution of 2 mm. Echo time was 90 ms and repetition time was 8,500 ms. QSIPrep 0.13.0RC1 (based on Nipype 1.6.0 Gorgolewski et al., 2011) was used for MRI pre-processing and reconstruction. Whole-brain white matter connectivity matrices (i.e. connectomes) were constructed for each child based on the Brainnetome atlas (Fan et al., 2016). For each pairwise combination of regions (N = 246), the number of streamlines intersecting them was estimated and transformed to a 246 x 246 streamline matrix.

Only a subsample of the children referred to CALM agreed to participate in the neuroimaging session (N = 248, Age scan = 10.92, SD = 2.07, 70% male). The neuroimaging cohort also included 77 children who formed a comparison group. These children (N = 77, Age scan =10.75, SD = 2.01) were

not referred by health and educational professionals but were recruited from the same schools attended by those who were referred. This non-referred / neurotypical group was included in the neuroimaging analyses, because unlike the cognitive and behavioural assessments, neuroimaging metrics do not allow for straightforward interpretation of age-typical levels of functioning.

Analysis Plan

Analyses were conducted in four steps: 1) missing data estimation; 2) self-organising map (SOM) and data-driven clustering; 3) comparison of profiles across executive function, neurodevelopmental, academic, and mental health functioning based on a series of chi-squares and t-tests; 4) pairwise comparisons and partial least squares discriminant analysis (PLSDA) to explore patterns of neural white matter organisation that distinguish the data-driven groups from the neurotypical participants who were included in the neuroimaging cohort. Details for each step of the analysis are outlined below.

Missing data

Missing data were estimated only for the assessments used in the SOMs. For all other assessments used in the validation analyses, missing data was not estimated to ensure it was independent and suitable for external validation. Missingness across executive function assessments used in the SOMs ranged from 0.7% to 25.3%. Data was complete for 59% of participants. Boys and girls were equally likely to have missing data ($\chi2 = 0.22$, $p = 0.64$) and so were diagnosed and undiagnosed participants ($\chi2 = 0.03$, $p = 0.87$). Missing data was estimated via the random forest nonparametric imputation procedure implemented in R package missForest (Stekhoven & Bühlmann, 2012). Participant sex and age were also added to the imputation model to improve estimates.

Self-organising maps (SOMs)

Self-organizing maps (SOMs) are a type of artificial neural network suitable for analysing high-dimensional data in a lower-dimensional space. SOMs use unsupervised learning to group similar input data together and create a topological map, where neighbouring nodes correspond to similar input patterns. Here we provide a brief conceptual overview of the implementation used in the current analysis, for more information about the SOM algorithm, see Wehrens & Buydens, (2007). SOMs

consist of a predefined number of nodes laid out on a grid – in this case, hexagonal nodes on 10 by 10 grid, where each node in the grid corresponded to a weight vector with the same dimensionality as the input data (N =16, corresponding to the number of executive function measures used to train the SOMs). The training of the map began with a subset of the data randomly assigned to the units. The process was repeated in 1000 iterations, and each time the weights of a best-matching unit (i.e., the node most alike the current training object based on the least shared Euclidean distance), and its neighbouring nodes, were updated to become more similar to the input data. In this way, neighbouring nodes became more alike. At the end of the training process, the weight vector for each node reflected the executive function scores of the children for whom that node was the best matching unit, with neighbouring nodes having similar weights. The resulting map represented a model of the multidimensional executive function data on which the SOM was trained, whereby children with similar profiles across the 16 executive function assessments "sat" closer in space. Age-standardised scores were used in these analyses. These were a combination of T-, scaled and standard scores, so they were converted to age-referenced z-scores to be on the same scale (score of 0 corresponded to age-typical performance, 1 = 1 SD above the age-expected mean, and -1 represented 1 SD below the age-expected mean). SOMs were trained using the *R* package kohonen (Wehrens & Buydens, 2007).

Data-driven clustering

The SOM maps the data in a continuous two-dimensional plane of nodes, where space indicates similarity. Therefore, carving the map into sections should yield groups with relatively homogenous profiles who are different from children sitting elsewhere on the map. Given the unsupervised nature of the SOMs, there is no pre-defined rationale for how many sections the map should be divided into. To identify the optimal number of groups, we used a data-driven method in which the nodal weight values from the SOM were submitted to a data-driven clustering. Following good practice recommendations, before clustering, the nodal weights were reduced using uniform manifold approximation and projection (Bathelt et al., 2021; Dalmaijer et al., 2022). The optimal number of clusters was chosen based on a consensus approach implemented in the R-package NbClust (Charrad et al., 2014). This method uses 30 indices for determining the number of clusters and converges on the best clustering scheme across

the different results obtained by combinations of the number of clusters, distance measures, and clustering methods. Once the nodes were grouped according to the similarity of their weights, we identified children assigned to each group of nodes. This provided us with clusters of children based on nodes they were assigned to in the original mapping. To check the clustering, each cluster distribution was plotted on the original map with the expectation that all cluster members ought to sit on neighbouring nodes within the original map.

Cluster characterisation & comparisons

To characterise the data-driven groups we compared their average scores on each of the assessments used in the SOM in a series of t-tests. False discovery correction was used to handle multiple comparisons. We also validated the data-driven clusters by testing whether group differences generalised to data not included in the SOM – in this case, learning assessments and self-reported mental health difficulties, neurodevelopmental diagnostic categories, and parent-reported dimensions of neurodiversity. For continuous data, t-tests were conducted using false discovery rate as a correction for multiple comparisons, and for categorical data, chi-square tests were used (e.g., to compare the number of children with a particular diagnosis in each subgroup relative to the overall distribution across the whole sample).

Neuroimaging comparisons

The neuroimaging analyses explored three levels of white matter connectome organisation: global, module, and regional hubs. For the first two, we used series of non-parametric comparisons contrasting the non-referred neurotypical group to the data-driven groups using false-discovery rate correction. This approach was favoured as it allowed us to explore both shared and subgroup-unique differences relative to the comparison group with a minimum number of comparisons. To explore differences in regional hub organisation we used a PLSDA (Brereton & Lloyd, 2014) , aiming to derive components of regional hubs that best explain group membership across the data-driven profiles and the non-referred neurotypical sample. Prior to analyses, the effects of age and in-scanner motion (average frame displacement) were regressed from each metric using a robust estimation approach.

At the global level, we focused on connectome efficiency and the average participation coefficient. Global efficiency describes the potential for information exchange in the connectome and is the average inverse distance from any one brain region to another (Sporns et al., 2007). The participation coefficient measures the proportion of possible connections each region has with regions from other modules. It is considered an index of modular segregation, with lower values indicating more segregation: regions with a high participation coefficient have strong connections to many modules, while regions with a low participation coefficient have strong connections to fewer modules (Baum et al., 2017). A global participation coefficient was derived by averaging across the 246 regions. These metrics were chosen because they have been previously linked to improved performance on tests of executive function in developing populations (Baum et al., 2017; Koenis et al., 2015).

To derive module organisation, we grouped the 246 regions of the Brainnetome atlas a priori into their corresponding functional networks as defined by Yeo et al. (2011). Subcortical nodes were also grouped together. This modular parcellation is based on independent functional data but has been previously shown to be a good representation of the organisation seen in the white matter connectome in developing populations (Baum et al., 2017). We then estimated modular strength for each one of the 7 networks and the subcortex. Significantly different results were followed-up to explore whether differences were driven by changes in within-module connectivity, between-module connectivity, or both.

Finally, PLSDA was chosen as a tool to explore links between regional connector hub strength and group membership. PLSDA is a statistical technique used to find relationships between two sets of variables: a predictor set and a response set. In our case these were connector hub strength and group membership respectively. The goal of PLSDA was therefore to identify the linear combination of connector hubs that best explains the variation in group membership. We focused on connector hubs because they play an important role in network specialisation across development and support the development of executive function (Baum et al., 2017; Jones et al., 2021; Zink et al., 2021). In accordance with prior work, connector hubs were defined as regions which were above the 70th percentile on both betweenness centrality and the participation coefficient (Jones et al., 2021). The

analysis was implemented in the R package mixOmics (Rohart et al., 2017). The model was evaluated by using 5-fold cross-validation repeated 50 times and the number of components to retain was chosen based the balanced error rate metric, accounting for the uneven number of participants across groups ($N_{cluster\ 1}$ = 95, $N_{cluster2}$ =74, $N_{cluster3}$ =79). The contribution of each hub (i.e., component loadings) to the PLSDA components was then evaluated using a bootstrap procedure (N= 1000), where a Procrustes rotation was applied to align the factors across iterations (Krishnan et al., 2011). This was done to test whether the bootstrapped confidence interval passed zero, and thus to establish which hubs reliably load on the PLSDA components. Group differences in component scores were compared using a non-parametric permutation procedure testing whether they significantly differed from chance when group labels were permuted 1000 times.

Results

SOMs and Clustering

Performance across all measures used in the SOM are presented in Table 1, and correlations between the measures in Figure 1. The outcomes of the trained SOM are presented in Figures 2, showing how the SOM represents the values for each weight vector (i.e., the weights that correspond to each individual executive function measure) across the grid of nodes. Each panel shows the distribution of weights for a different executive function measure. Data-driven consensus clustering applied to the SOM suggested that a three-cluster solution was optimal (the map separated by cluster can be seen in Figure 3).

Cluster Characterisation

The group means for each of the three data-driven groups on the executive function assessments used in the SOM are shown in Figure 4. One group (Cluster 1, N = 204) was characterised by close to age-expected levels across the rating-based assessments, with fewer difficulties compared to the other two clusters. Their performance on the task-based measures was poorer than their ratings, and below age-expected levels. Overall, they had better task-based performance compared to children in Cluster 2 and poorer performance than those in Cluster 3. Based on the discrepancy between their rating- and task-

based performance, children in Cluster 1 are referred to as having an inconsistent profile of predominantly task-based difficulties.

A second group, Cluster 2 (N=187), was characterised by widespread difficulties across both the rating- and task-based executive function assessments and was therefore labelled as having consistent difficulties. The third group, Cluster 3 (N = 175), included children with the opposite profile to those in Cluster 1. They had an inconsistent profile with close to age-expected levels of performance on the task-based measures with pronounced difficulties on the rating-based assessments that were indistinguishable from those of the children in Cluster 2 who had a profile of consistent difficulties. Children in Cluster 3 are referred to as having an inconsistent profile of predominantly rating-based difficulties.

There were significant differences in the average age of children in each of the three Clusters ($F = 9.56$, $p < .01$): children in Cluster 1 who had an inconsistent profile of predominantly task-based difficulties were on average younger than children in Cluster 2 who had a consistent profile of difficulties ($t = -3.50$, $p < .001$, mean difference = 0.71 years) and children in Cluster 3 who had predominantly rating-based difficulties ($t = -3.06$, $p < .002$, mean difference = 0.62 years). There was no significant difference in age between children in Clusters 2 and 3. To explore whether child sex was related to group membership, while accounting for the overrepresentation of boys in the sample as a whole, a series of chi-square tests were conducted to compare the sex distribution observed in each cluster to the sex distribution of the whole sample (Figure 4). Girls were overrepresented in Cluster 1 (profile with predominantly task-based difficulties, $\chi^2 = 5.66$, $p = .02$), and underrepresented in Cluster 3 (profile with predominantly rating-based difficulties, $\chi^2 = 7.94$, $p = .005$). The distribution of boys and girls in Cluster 2 (consistent difficulties group) matched that of the whole sample ($\chi^2 = 0.06$, $p = .81$), suggesting that boys and girls were equally likely to have this profile.

Cluster validation

To validate differences between the three clusters, their performance was compared across measures of learning, mental health, and dimensions of neurodiversity, and the distribution of children with

diagnosed neurodevelopmental conditions in each cluster was compared to the distribution across the whole sample. Children in Cluster 3 (inconsistent profile of predominantly rating-based difficulties) had significantly better performance across measures of reading and maths relative to children in the other two clusters (Figure 5). Those in Cluster 2 (consistent difficulties profile) had significantly lower performance on the maths measure relative to children in Clusters 1 and 3, but their reading performance was indistinguishable from that of children in Cluster 1 who had predominantly task-based difficulties. For mental health, children in Cluster 3 (predominantly rating-based difficulties) had significantly elevated levels of anxiety and depression relative to children in the other two clusters; their scores were not significantly different from one another (see Figure 5).

Autistic participants and those with an ADHD diagnosis were underrepresented in Cluster 1, the group with predominately task-based difficulties (ADHD: χ^2= 39.73, p<.001; Autism: χ^2= 12.11, p<.0001), but were represented in similar proportions in the other two clusters. It should be noted that 41% of autistic participants also had an ADHD diagnosis, meaning separate conclusions cannot be drawn for the two conditions. Children with a dyslexia diagnosis were present in each cluster, and there was no relationship between dyslexia diagnosis and group membership (ps >.05). A similar pattern emerged across dimensions associated with neurodevelopmental conditions. Across the five dimensional measures (inattention, hyperactivity/impulsivity, general communication skills, social communication and interests), children in Cluster 1 (predominantly task-based difficulties) had significantly fewer difficulties than those in Clusters 2 and 3, who did not differ significantly from one another. In other words, there was no clear one-to-one correspondence between data-driven group membership and neurodevelopmental diagnostic dimensions. Instead, relative to the group with primarily task-based difficulties (Cluster 1), the groups with consistent and predominantly rating-based difficulties (Clusters 2 and 3) included more children with diagnosed neurodevelopmental conditions and experienced more neurodevelopmental difficulties on dimensional measures.

Neuroimaging comparisons

At the global level (Figure 6), children with consistent difficulties across rating- and task-based measures (Cluster 2) and those with mainly rating-based difficulties (Cluster 3) had significantly lower

global connectome efficiency compared to the non-referred neurotypical group. The three data-driven groups did not differ significantly from the non-referred comparison group in average connectome participation coefficient. Looking at modular organisation, Clusters 2 (consistent difficulties) and 3 (predominantly rating-based difficulties) had similar patterns of neural organisation that differentiated them from the non-referred comparison group: they both showed reduced modular strength in the limbic network and the subcortex (Figure S1). The group with primarily task-based difficulties (Cluster 1) showed the same pattern of reduced modular strength in the limbic network and the subcortex but was also characterised by increased strength in somatomotor and ventral attention modules relative to the comparison sample (Figure S1). Follow-up analysis suggested that all these differences were driven by differences in between - rather than within - modular connection strength (Figure 6).

Next, we identified 22 connector hubs that included regions within middle frontal gyrus, orbital gyrus, precentral gyrus, insula, cingulate gyrus, and the subcortex. The cross-validated balanced error rate across all models explored (k =1:5) suggested that the most accurate PLSDA model was the one with two components. The bootstrapped hub loadings onto the PLSDA components, which had 95% confidence intervals that did not cross zero, are shown in Figure 7. For the first component, robust loadings included regions mostly within the insula, as well as within the dorsolateral prefrontal cortex, superior temporal gyrus, and the basal ganglia. Permutation analyses suggested that relative to the non-referred comparison group, children with an inconsistent profile of predominantly task-based difficulties (Cluster 1) scored significantly lower on this component (p=.003). For the second component, all loadings that were shown to be robust in the bootstrapping procedure were frontal and parietal subregions. The second component distinguished participants with a profile of consistent difficulties across the executive function measures (Cluster 2) from the non-referred comparison group (p = .02), with component scores significantly higher in the comparison group.

Discussion

This study used an unsupervised machine learning approach to map profiles of executive function among a large neurodivergent sample of children. Topographical maps were used to represent the profiles, which were subsequently carved into clusters representing homogeneous groups of children

with similar profiles of executive function. Three clusters were identified, each with a distinct profile of executive function. The clusters could also be distinguished on measures of learning and mental health, but they cut across traditional neurodevelopmental diagnostic taxonomies: for the most part, they did not correspond to diagnostic categories or to dimensions of neurodiversity associated with common conditions such as ADHD or autism.

Three profiles of executive function were identified. One was characterised by difficulties that were consistent across the task- and rating-based measures of executive function (Cluster 2). The other two profiles were inconsistent across the two measurement types: children in Cluster 1 had close to age-expected levels on the rating-based measures but had difficulties on the performance-based tasks, while children in Cluster 3 had the opposite profile of close to age-expected performance on the task-based measures with pronounced difficulties on the rating-based assessments. In total, more than half the sample had an inconsistent profile. This provides further evidence that the two types of assessment (ratings and tasks) are dissociable, consistent with previous studies that have used latent variable or predictive validity approaches (e.g., Gerst et al., 2017; Nin et al., 2022; Soto et al., 2020; Tamm & Peugh, 2019; Toplak et al., 2013). The finding that a substantial proportion of our transdiagnostic neurodivergent sample showed little overlap between performance and rating based measures is also consistent with previous neurodevelopmental literature, which has shown that the two domains of executive function can be selectively affected within the boundaries of a single neurodevelopmental condition (e.g., those with ADHD, Biederman et al., 2008). Overall, the current findings support the idea that two types of measurement provide non-redundant information and should not be used interchangeably in the context of neurodevelopment.

Cluster comparisons: behaviour

Children in Cluster 1, who were characterised by difficulties on the cognitive performance-based tasks with near to age-typical everyday executive function behaviours, were significantly younger than children in other two clusters, were less likely to have a diagnosed neurodevelopmental condition, and had fewer difficulties on dimensions of neurodiversity associated with ADHD, developmental language disorder and autism than children in the other two clusters. Girls were under-represented in

this cluster. Children in this cluster performed more poorly on both learning measures than children in Cluster 3, who had the opposite profile of better task- than rating-based performance, and their maths scores were equivalent to those in Cluster 2 who had a consistent but poor executive function profile.

The broad profile of children in Cluster 1 suggests that diagnoses are less common among younger children, girls, and those with everyday behaviours that appear to be age-typical, despite poor performance on cognitive task-based measures of executive function and learning. This might reflect diagnostic systems that rely on observations of behaviour rather than performance on cognitive tests (the everyday behaviours of this group were almost age-typical), and in which children are not typically formally assessed for conditions such as ADHD until they are 7 years or older (American Psychiatric Association, 2013; Hoang et al., 2019; Root et al., 2019). Lower rates of diagnoses among this group, in which girls were over-represented, could also reflect differences in the outward characteristics of neurodivergent girls and boys: girls are more likely to use behavioural camouflaging strategies, appearing more able from others' perspectives (Dean et al., 2017; Hiller et al., 2014; Hull et al., 2020), meaning they are less likely to receive a diagnosis (Dhuey & Lipscomb, 2010; Lockwood Estrin et al., 2021). Patterns of difficulties captured by parent/carer ratings on dimensions of neurodiversity associated with neurodevelopmental difficultes (e.g. ratings of inattention, restricted social interests etc.) reflected the same pattern as the diagnostic categories: fewer difficulties were present in children in this cluster relative to theose in the other clusters. This could again reflect masking / camouflaging, as well as socially constructed gender-biased or stereotypical views of boys as being disruptive (e.g., Sciutto et al., 2004). Together these findings underscore the need to include cognitive task-based assessments and female-representative characteristics in systems designed to identify children's additional needs (see Guy et al., 2022 further discussion of this issue).

Children with the opposite profile of executive function, those in Cluster 3 with close to age-typical task-based performance and greater difficulties on the rating-based assessments, had significantly better performance on the learning measures than children in the other two clusters. This is consistent with a vast literature showing strong associations between performance on laboratory-based assessments of executive function and learning (Bull et al., 2008; Holmes et al., 2020; St Clair-

Thompson & Gathercole, 2006). They also had significantly elevated self-reported internalising difficulties (anxiety and depression) compared to children in the other two clusters. This group also self-reported more internalising difficulties relative to the other two clusters, potentially suggesting a link between higher levels of executive function rating difficulties and mental health. Notably, this could not be entirely attributed to rater biases, given that mental health was self-rated and executive function ratings were provided by parents. The relationship between mental health difficulties and a profile of inconsistent and predominantly rating-based difficulties has previously been observed using different methods in this cohort (Williams et al., 2022). Others have suggested that rating-based difficulties occurring in the absence of any task-based performance problems arise through emotional rather than cognitive mechanisms (Vaidya et al., 2020; Williams et al., 2022), but longitudinal and/or causal evidence is needed to test this hypothesis.

Finally, those with the consistent profile of executive function difficulties had below-age-expected performance across both types of executive function assessments. Relative to the other clusters, they had similar ratings of executive function difficulties to children in Cluster 3, but significantly poorer performance on the task-based assessments than children in both other clusters. Their reading and mental health ratings were equivalent to those in Cluster 1, but their maths performance was significantly. Overall, they showed commonalities and differences with children in the other two clusters, which were not entirely attributable to measurement type (e.g., even though their rating-based executive function difficulties were equivalent to those in Cluster 3, they had significantly different mental health ratings).

Cluster comparisons: neuroimaging

We observed both shared and cluster-specific patterns of neural white matter organisation differentiating data-driven groups from the comparison sample of neurotypical children who had not been referred for attention, learning, and/or memory difficulties. Children with either a consistent profile of executive difficulties (Cluster 2) and those with predominantly task-based difficulties (Cluster 3) had decreased global connectome efficiency relative to the comparison sample, but no group

differences were observed in global participation. This was unexpected given previous reports of an association between modular segregation and executive function (Baum et al., 2017).

Looking at modular organisation, all three data-driven groups showed reduced connection strength within the limbic network and the subcortex relative to the non-referred neurotypical comparison group, with follow-up analyses showing that differences were driven by reduced inter-module connections. This is consistent with the idea that limbic and subcortical areas play a role in goal-directed cognition, and that differences in this circuitry are related to executive functioning in neurodevelopmental conditions (e.g., Arnsten & Rubia, 2012). The group with predominantly task-based difficulties (Cluster 1) also showed increased inter-module connection strength in the somatomotor and ventral attention modules, a pattern that was also observed in the PLSDA: the component of connector hubs differentiating this group from the non-referred neurotypical comparison group involved multiple insular regions of the somatomotor and ventral attention modules. Somatomotor and ventral attention networks undergo substantial restructuring over the course of development, becoming more structurally segregated with age (Baum et al., 2017; Grayson & Fair, 2017). Observing differences in this network among children with predominantly task-based difficulties is consistent with their implicated role in task-based executive functions (Reineberg et al., 2015). Additionally, several of the hubs which loaded robustly on the PLSDA component that differentiated this group from the neurotypical sample shared perception as a behavioural metadata label within the BrainMap Database (www.brainmap.org/ taxonomy).

Children with a profile of consistent difficulties across the task- and rating-based measures (Cluster 2) were uniquely distinguished from the non-referred comparison group through a component of frontal and parietal connector hubs identified through the PLSDA. This is consistent with decades of research that has established a role for frontal and parietal regions in supporting executive function (see Friedman & Robbins, 2022). Exploratory cross-checking of the behavioural metadata labels for the hubs robustly loading on this component within the BrainMap Database suggested that many of them were functionally implicated in action execution.

Finally, considering other similarities, the two components identified by the PLSDA that differentiated the two groups with task-based executive function difficulties (Cluster 1 and 2) from the non-referred comparison sample had a single shared robust loading. This region, located in the right dorsal area 9/46 within the middle frontal gyrus, is functionally associated with task-based executive function performance on measures of working memory and attention (e.g., Jung et al., 2022).

Limitations

The measurement types used were limited by those available in the cohort data. As such, it is possible, for example that stronger links between the everyday manifestations of executive function difficulties captured by the rating scales and learning could have been detected if we did not have to rely on measures of learning administered under optimal testing conditions, but instead had access to school-based measures of academic achievement that reflect learning and assessment in everyday conditions. Methodologically, the combination of a continuous multidimensional mapping method with a data-driven clustering algorithm means that inevitably some children will sit close to the cluster boundaries within the map. In terms of the neuroimaging results, it is important to acknowledge that the use of multimodal neuroimaging and larger samples may reveal further and equally important differences across groups, which we consider an important future direction. Finally, the exploratory nature of this investigation should be noted. We encourage further efforts to replicate and extend these findings to better understand the diversity of executive function profiles among neurodivergent individuals.

Conclusion

In summary, we used a machine learning approach to map the executive function profiles of large neurodevelopmentally neurodivergent sample. We identified three distinct profiles that were validated by differences across measures of learning, mental health, and neural white matter organisation. These data add to the growing evidence base for two conceptualisations of executive function: one that reflects higher-order cognitive skills and can be measured by cognitive tasks (Miyake et al., 2000), and the other that reflects the application of cognitive control in everyday contexts and can be measured by rating scales (Doebel, 2020). The majority of our neurodivergent sample had inconsistent profiles across the two domains, suggesting that researchers and practitioners should not use these two types of

assessments interchangeably in the context of neurodevelopment. They should instead be used together in a complementary manner.

References

Alloway, T. P. (2007). *Automated Working Memory Assessment: Manual*. Pearson.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The working memory rating scale: A classroom-based behavioral assessment of working memory. *Learning and Individual Differences*, *19*(2), 242–245. https://doi.org/10.1016/j.lindif.2008.10.003

Ameis, S. H., Lerch, J. P., Taylor, M. J., Lee, W., Viviano, J. D., Pipitone, J., Nazeri, A., Croarkin, P. E., Voineskos, A. N., Lai, M. C., Crosbie, J., Brian, J., Soreni, N., Schachar, R., Szatmari, P., Arnold, P. D., & Anagnostou, E. (2016). A diffusion tensor imaging studyin children with ADHD, autism spectrum disorder, OCD, and matched controls: Distinct and non-distinct white matter disruption and dimensional brain-behavior relationships. *American Journal of Psychiatry*, *173*(12), 1213–1222. https://doi.org/10.1176/appi.ajp.2016.15111435

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub.

Arnsten, A. F. T., & Rubia, K. (2012). Neurobiological circuits regulating attention, cognitive control, motivation, and emotion: Disruptions in neurodevelopmental psychiatric disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(4), 356–367. https://doi.org/10.1016/j.jaac.2012.01.008

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2021). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/jcpp.13481

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*, *25*(3), 157–173. https://doi.org/10.1093/arclin/acq014

Barkley, R. A., & Murphy, K. R. (2011). The Nature of Executive Function (EF) Deficits in Daily Life Activities in Adults with ADHD and Their Relationship to Performance on EF Tests. *Journal of Psychopathology and Behavioral Assessment*, *33*(2), 137–158. https://doi.org/10.1007/s10862-011-9217-x

Bathelt, J., Holmes, J., Astle, D. E., Holmes, J., Gathercole, S., Astle, D., Manly, T., & Kievit, R. (2018). Data-Driven Subtyping of Executive Function–Related Behavioral Problems in Children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *57*(4), 252-262.e4. https://doi.org/10.1016/j.jaac.2018.01.014

Bathelt, J., Vignoles, A., & Astle, D. E. (2021). Just a phase? Mapping the transition of behavioural problems from childhood to adolescence. *Social Psychiatry and Psychiatric Epidemiology*, *56*(5), 821–836. https://doi.org/10.1007/s00127-020-02014-4

Baum, G. L., Ciric, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., Kahn, A. E., Vandekar, S. N., Rupert, P. E., Quarmley, M., Cook, P. A., Elliott, M. A., Ruparel, K., Gur, R. E., Gur, R. C., Bassett, D. S., & Satterthwaite, T. D. (2017). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth. *Current Biology*, *27*(11), 1561-1572.e8. https://doi.org/10.1016/j.cub.2017.04.051

Beare, R., Adamson, C., Bellgrove, M. A., Vilgis, V., Vance, A., Seal, M. L., & Silk, T. J. (2017). Altered structural connectivity in ADHD: a network based analysis. *Brain Imaging and Behavior*, *11*(3), 846–858. https://doi.org/10.1007/s11682-016-9559-9

Benallie, K. J., McClain, M. B., Bakner, K. E., Roanhorse, T., & Ha, J. (2021). Executive functioning in children with ASD + ADHD and ASD + ID: A systematic review. *Research in Autism Spectrum Disorders*, *86*, 101807. https://doi.org/10.1016/J.RASD.2021.101807

Biederman, J., Petty, C. R., Fried, R., Black, S., Faneuil, A., Doyle, A. E., Seidman, L. J., & Faraone, S. V. (2008). Discordance Between Psychometric Testing and Questionnaire-Based Definitions of Executive Function Deficits in Individuals With ADHD. *Journal of Attention Disorders*, *12*(1), 92–102. https://doi.org/10.1177/1087054707305111

Bishop, D. V. M. (2003). *Children's Communication Checklist-2 (CCC-2)*. Pearson Assessment.

Bloemen, A. J. P., Oldehinkel, A. J., Laceulle, O. M., Ormel, J., Rommelse, N. N. J., & Hartman, C. A. (2018). *Psychological Medicine The association between executive functioning and psychopathology: General or specific?* https://doi.org/10.1017/S0033291717003269

Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics*, *28*(4), 213–225. https://doi.org/10.1002/cem.2609

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-Term Memory, Working Memory, and Executive Functioning in Preschoolers: Longitudinal Predictors of Mathematical Achievement at Age 7 Years. *Developmental Neuropsychology*, *33*(3), 205–228. https://doi.org/10.1080/87565640801982312

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*, 1–36. https://doi.org/10.18637/jss.v061.i06

Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy*, *43*(3), 309–322. https://doi.org/10.1016/J.BRAT.2004.02.004

Conners, C. K. (1992). *Conners' Continuous Performance Test user's manual*. Multi-Health System.

Conners, C. K. (2008). *Conners 3rd edition (Conners 3)* (Third Edit). Multi-Health Systems Toronto, Ontario, Canada.

Dalmaijer, E. S., Nord, C. L., & Astle, D. E. (2022). Statistical power for cluster analysis. *BMC Bioinformatics*, *23*(1), 1–28. https://doi.org/10.1186/s12859-022-04675-1

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007

Dean, M., Harwood, R., & Kasari, C. (2017). The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. *Autism*, *21*(6), 678–689. https://doi.org/10.1177/1362361316671845

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system (D-KEFS)*. The Psychological Corporation.

Dhuey, E., & Lipscomb, S. (2010). Disabled or young? Relative age and special education diagnoses in schools. *Economics of Education Review*, *29*(5), 857–872. https://doi.org/10.1016/j.econedurev.2010.03.006

Doebel, S. (2020). Rethinking Executive Function and Its Development. *Perspectives on Psychological Science*, *15*(4), 942–956. https://doi.org/10.1177/1745691620904771

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-10301-1

Ellingson, J. M., Corley, R., Hewitt, J. K., & Friedman, N. P. (2019). A prospective study of alcohol involvement and the dual-systems model of adolescent risk-taking during late adolescence and emerging adulthood. *Addiction*, *114*(4), 653–661. https://doi.org/10.1111/add.14489

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., & Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex*, *26*(8), 3508–3526. https://doi.org/10.1093/cercor/bhw157

Friedman, N. P., & Gustavson, D. E. (2022). Do Rating and Task Measures of Control Abilities Assess the Same Thing? *Current Directions in Psychological Science*, *31*(3), 262–271. https://doi.org/10.1177/09637214221091824

Friedman, N. P., Hatoum, A. S., Gustavson, D. E., Corley, R. P., Hewitt, J. K., & Young, S. E. (2020). Executive Functions and Impulsivity Are Genetically Distinct and Independently Predict Psychopathology: Results From Two Adult Twin Studies. *Clinical Psychological Science*, *8*(3), 519–538. https://doi.org/10.1177/2167702619898814

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/J.CORTEX.2016.04.023

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, *47*(1), Article 1. https://doi.org/10.1038/s41386-021-01132-0

Gerst, E. H., Cirino, P. T., Fletcher, J. M., & Yoshida, H. (2017). Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *23*(4), 381–407. https://doi.org/10.1080/09297049.2015.1120860

Geurts, H. M., Verté, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and Psychiatry*, *45*(4), 836–854. https://doi.org/10.1111/j.1469-7610.2004.00276.x

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior rating inventory of executive function: BRIEF*. Psychological Assessment Resources Odessa, FL.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*(August). https://doi.org/10.3389/fninf.2011.00013

Grayson, D. S., & Fair, D. A. (2017). Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. *NeuroImage*, *160*, 15–31. https://doi.org/10.1016/j.neuroimage.2017.01.079

Guy, J., Mareva, S., Franckel, G., Team, the C., & Holmes, J. (2022). Dimensions of cognition, behaviour, and mental health in struggling learners: A spotlight on girls. *JCPP Advances*, *2*(4), e12082. https://doi.org/10.1002/jcv2.12082

Hiller, R. M., Young, R. L., & Weber, N. (2014). Sex Differences in Autism Spectrum Disorder based on DSM-5 Criteria: Evidence from Clinician and Teacher Reporting. *Journal of Abnormal Child Psychology*, *42*(8), 1381–1393. https://doi.org/10.1007/s10802-014-9881-x

Hoang, U., James, A. C., Liyanage, H., Jones, S., Joy, M., Blair, M., Rigby, M., & Lusignan, S. de. (2019). Determinants of inter-practice variation in ADHD diagnosis and stimulant prescribing: Cross-sectional database study of a national surveillance network. *BMJ Evidence-Based Medicine*, *24*(4), 155–161. https://doi.org/10.1136/bmjebm-2018-111133

Holmes, J., Bryant, A., & Gathercole, S. E. (2019). Protocol for a transdiagnostic study of children with problems of attention, learning and memory (CALM). *BMC Pediatrics*, *19*(1), 10. https://doi.org/10.1186/s12887-018-1385-3

Holmes, J., Guy, J., Kievit, R. A., Bryant, A., Mareva, S., Team, the C., & Gathercole, S. E. (2020). Cognitive dimensions of learning in children with problems in attention, learning and memory. *Journal of Educational Psychology*. http://dx.doi.org/10.1037/edu0000644

Holmes, J., Hilton, K. A., Place, M., Alloway, T. P., Elliott, J. G., & Gathercole, S. E. (2014). Children with low working memory and children with ADHD: same or different? *Frontiers in Human Neuroscience*, *8*, 976. https://doi.org/10.3389/fnhum.2014.00976

Hull, L., Petrides, K. V., & Mandy, W. (2020). The Female Autism Phenotype and Camouflaging: A Narrative Review. *Review Journal of Autism and Developmental Disorders*, *7*(4), 306–317. https://doi.org/10.1007/s40489-020-00197-9

Jones, J. S., the CALM Team, & Astle, D. E. (2021). A transdiagnostic data-driven study of children's behaviour and the functional connectome. *Developmental Cognitive Neuroscience*, *52*, 101027. https://doi.org/10.1016/J.DCN.2021.101027

Jung, J., Lambon Ralph, M. A., & Jackson, R. L. (2022). Subregions of DLPFC Display Graded yet Distinct Structural and Functional Connectivity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(15), 3241–3252. https://doi.org/10.1523/JNEUROSCI.1216-21.2022

Koenis, M. M. G., Brouwer, R. M., van den Heuvel, M. P., Mandl, R. C. W., van Soelen, I. L. C., Kahn, R. S., Boomsma, D. I., & Hulshoff Pol, H. E. (2015). Development of the brain's structural network efficiency in early adolescence: A longitudinal DTI twin study. *Human Brain Mapping*, *36*(12), 4938–4953. https://doi.org/10.1002/hbm.22988

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, *56*(2), 455–475. https://doi.org/10.1016/j.neuroimage.2010.07.034

Lockwood Estrin, G., Milner, V., Spain, D., Happé, F., & Colvert, E. (2021). Barriers to Autism Spectrum Disorder Diagnosis for Young Women and Girls: A Systematic Review. *Review Journal of Autism and Developmental Disorders*, *8*(4), 454–470. https://doi.org/10.1007/s40489-020-00225-8

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology*, *32*(6), 643–654. https://doi.org/10.1093/jpepsy/jsl054

Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P., & Tucker-Drob, E. M. (2018). 'Same But Different': Associations Between Multiple Aspects of Self-Regulation, Cognition, and Academic Abilities. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000224

Manly, T., Anderson, V., Crawford, G., George, M., & Robertson, I. (2016). *Test of Everyday Attention for Children, Second Edition (TEA-Ch2)*. Pearson. https://www.pearsonclinical.co.uk/store/ukassessments/en/adaptive-behavior/Test-of-Everyday-Attention-for-Children%2C-Second-Edition/p/P100009285.html

McClain, M. B., Golson, M. E., & Murphy, L. E. (2022). Executive functioning skills in early childhood children with autism, intellectual disability, and co-occurring autism and intellectual disability. *Research in Developmental Disabilities*, *122*, 104169. https://doi.org/10.1016/j.ridd.2021.104169

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. https://doi.org/10.1006/cogp.1999.0734

Nin, V., Delgado, H., Muniz-Terrera, G., & Carboni, A. (2022). Partial agreement between task and BRIEF-P-based EF measures depends on school socioeconomic status. *Developmental Science*, *25*(5), e13241. https://doi.org/10.1111/desc.13241

Peng, P., & Fuchs, D. (2016). A Meta-Analysis of Working Memory Deficits in Children With Learning Difficulties: Is There a Difference Between Verbal Domain and Numerical Domain? *Journal of Learning Disabilities*, *49*(1), 3–20. https://doi.org/10.1177/0022219414521667

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage*, *104*, 69–78. https://doi.org/10.1016/j.neuroimage.2014.09.045

Rohart, F., Gautier, B., Singh, A., & Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Root, A., Brown, J. P., Forbes, H. J., Bhaskaran, K., Hayes, J., Smeeth, L., & Douglas, I. J. (2019). Association of Relative Age in the School Year With Diagnosis of Intellectual Disability, Attention-Deficit/Hyperactivity Disorder, and Depression. *JAMA Pediatrics*, *173*(11), 1068–1075. https://doi.org/10.1001/jamapediatrics.2019.3194

Sciutto, M. J., Nolfi, C. J., & Bluhm, C. (2004). Effects of Child Gender and Symptom Type on Referrals for ADHD by Elementary School Teachers. *Journal of Emotional and Behavioral Disorders*, *12*(4), 247–253. https://doi.org/10.1177/10634266040120040501

Shallice, T. (1982). Specifc impairments of planning. *Philosophical Transactions of the Royal Society of London, Biology*, *289*, 199–209.

Snyder, H. R., Friedman, N. P., & Hankin, B. L. (2021). Associations Between Task Performance and Self-Report Measures of Cognitive Control: Shared Versus Distinct Abilities. *Assessment*, *28*(4), 1080–1096. https://doi.org/10.1177/1073191120965694

Soto, E. F., Irwin, L. N., Chan, E. S. M., Spiegel, J. A., & Kofler, M. J. (2021). Executive functions and writing skills in children with and without ADHD. *Neuropsychology*, *35*, 792–808. https://doi.org/10.1037/neu0000769

Soto, E. F., Kofler, M. J., Singh, L. J., Wells, E. L., Irwin, L. N., Groves, N. B., & Miller, C. E. (2020). Executive functioning rating scales: Ecologically valid or construct invalid? *Neuropsychology*, *34*(6), 605–619. https://doi.org/10.1037/neu0000681

Sporns, O., Honey, C., & Kötter, R. (2007). Identification and Classification of Hubs in Brain Networks. *PloS One*, *2*(10), e1049. https://doi.org/10.1371/journal.pone.0001049

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*(4), 745–759. https://doi.org/10.1080/17470210500162854

Stekhoven, D. J., & Bühlmann, P. (2012). Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Tamm, L., & Peugh, J. (2019). Concordance of teacher-rated and performance-based measures of executive functioning in preschoolers. *Child Neuropsychology*, *25*(3), 410–424. https://doi.org/10.1080/09297049.2018.1484085

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Vaidya, C. J., You, X., Mostofsky, S., Pereira, F., Berl, M. M., & Kenworthy, L. (2020). Data-driven identification of subtypes of executive function across typical development, attention deficit hyperactivity disorder, and autism spectrum disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *61*(1), 51–61. https://doi.org/10.1111/jcpp.13114

Wechsler, D. (2005). *Wechsler Individual Achievement Test II* (Second Edi). Pearson Assessment.

Wehrens, R., & Buydens, L. M. C. (2007). Self- and Super-organizing Maps in R: The kohonen Package. *Journal of Statistical Software*, *21*, 1–19. https://doi.org/10.18637/jss.v021.i05

Wierenga, L. M., van den Heuvel, M. P., Oranje, B., Giedd, J. N., Durston, S., Peper, J. S., Brown, T. T., Crone, E. A., & The Pediatric Longitudinal Imaging, Neurocognition, and Genetics Study. (2018). A multisample study of longitudinal changes in brain network architecture in 4–13-year-old children. *Human Brain Mapping*, *39*(1), 157–170. https://doi.org/10.1002/hbm.23833

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/ hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*(11), 1336–1346. https://doi.org/10.1016/j.biopsych.2005.02.006

Williams, K. L., Holmes, J., Farina, F., Vedechkina, M., Bennett, M. P., & The CALM team. (2022). Inconsistencies between Subjective Reports of Cognitive Difficulties and Performance on Cognitive Tests are Associated with Elevated Internalising and Externalising Symptoms in Children with Learning-related Problems. *Research on Child and Adolescent Psychopathology*, *50*(12), 1557–1572. https://doi.org/10.1007/s10802-022-00930-4

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1–9. https://doi.org/10.1016/J.LINDIF.2012.10.004

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fisch, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Zink, N., Lenartowicz, A., & Markett, S. (2021). A new era for executive function research: On the transition from centralized to distributed executive functioning. *Neuroscience & Biobehavioral Reviews*, *124*, 235–244. https://doi.org/10.1016/j.neubiorev.2021.02.011

Acknowledgements

The CALM Team includes lead investigators Duncan Astle, Kate Baker, Susan Gathercole, Joni Holmes, Rogier Kievit and Tom Manly. Data collection is assisted by a team of researchers and PhD students that includes Danyal Akarca, Joe Bathelt, Madalena Bettencourt, Marc Bennett, Giacomo Bignardi, Sarah Bishop, Erica Bottacin, Lara Bridge, Diandra Brkic, Annie Bryant, Sally Butterfield, Elizabeth Byrne, Gemma Crickmore, Edwin Dalmaijer, Fanchea Daly, Tina Emery, Laura Forde, Grace Franckel, Delia Furhmann, Andrew Gadie, Sara Gharooni, Jacalyn Guy, Erin Hawkins, Agnieszka Jaroslawska, Sara Joeghan, Amy Johnson, Jonathan Jones, Rebeca Ianov-Vitanov, Christian Iordanov, Silvana Mareva, Jessica Martin, Alicja Monaghan, Elise Ng-Cordell, Sinead O'Brien, Cliodhna O'Leary, Joseph Rennie, Andrea Santangelo, Ivan Simpson-Kent, Roma Siugzdaite, Tess Smith, Stephani Uh, Maria Vedechkina, Francesca Woolgar, Natalia Zdorovtsova, Mengya Zhang. The authors wish to thank the many professionals working in children's services in the South-East and East of England for their support, and to the children and their families for giving up their time to visit the clinic. The CALM dataset is not yet available as the study is still ongoing. The data will be made available via managed open access once the study is complete.

S.M. and J.H. conceived and designed the study, analysed the data, and wrote and edited the manuscript. The CALM team collected the data. Dr Danyal Akaca pre-processed the neuroimaging data. All authors approved the final version. Ethical approval was granted by the National Health Service. Parents/caregivers provided written consent and child verbal assent was obtained. The authors have declared that they have no competing or potential conflicts of interest.

Table 1

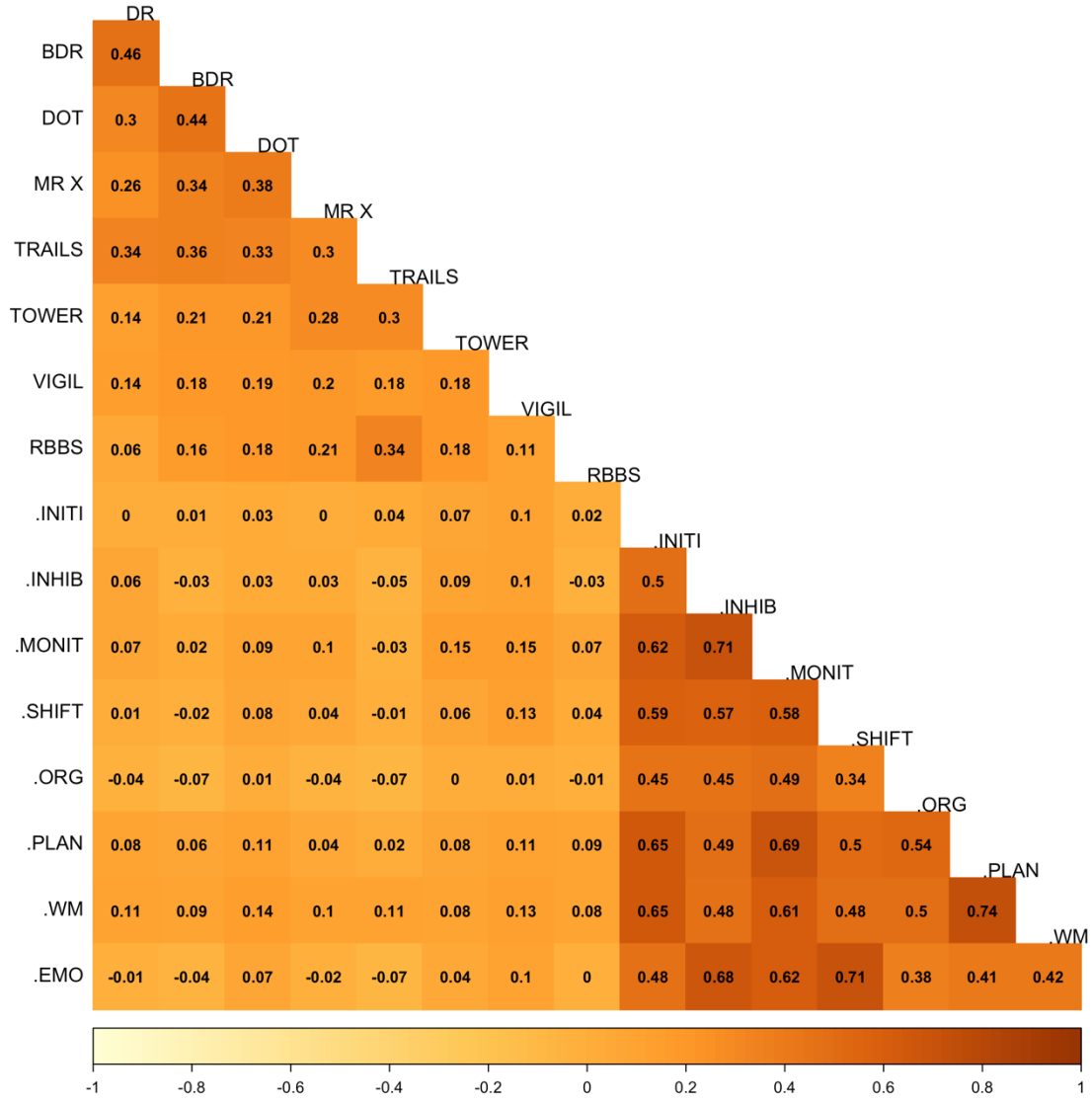
Norm-referenced age-corrected means and standard deviations for the whole sample across all measures used in the self-organising map and clustering.

	Mean	SD
AWMA: Digit recall	91.62	15.59
AWMA: Backward digit recall	91.39	11.46
AWMA: Dot matrix	90.02	15.60
AWMA: Mr X	95.90	14.43
DKEFS: Trails: number-letter sequencing	6.14	3.44
DKEFS: Tower	9.41	2.28
TEA-Ch2: Vigil	8.26	3.20
TEA-Ch2: Reds, Blues, Bags and Shoes	7.69	3.36
BRIEF: Initiation	66.46	10.43
BRIEF: Inhibition	66.16	15.15
BRIEF: Monitoring	65.51	10.71
BRIEF: Shifting	68.24	14.52
BRIEF: Organisation	60.08	9.82
BRIEF: Planning	70.41	9.57
BRIEF: Working memory	73.28	9.60
BRIEF: Emotional control	64.60	13.30

Note. AWMA = Automated Working Memory Assessment; DKEFS = Delis-Kaplan Executive Function Battery; TEA-Ch2 = Test of Everyday Attention for Children 2; BRIEF = Behaviour Rating Inventory of Executive Function.

Figure 1

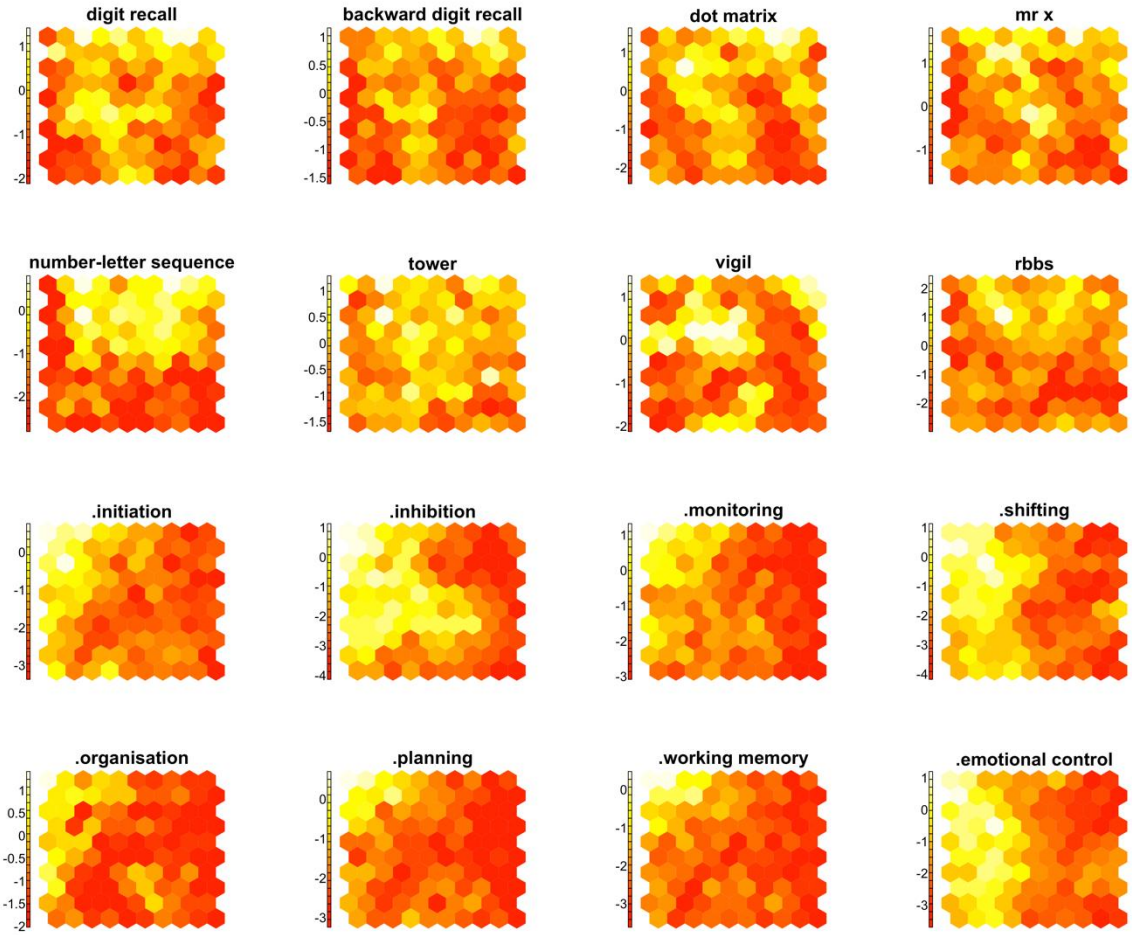
Pearson correlations across all norm-referenced age-corrected executive function assessment scores used in the self-organising map



Note. DR = AWMA Digit recall, BRD= AWMA Backward digit recall, DOT = AWMA Dot matrix, MR X = AWMA Mr X, TRAILS = DKEFS number-letter sequencing, TOWER = DKEFS Tower, VIGIL = TEA-Ch2 Vigil, SWITCH = TEA-Ch2 Reds, Blues, Bags and Shoes, .INITI = BRIEF Initiation, INHIB =BRIEF Inhibition, .MONIT = BRIEF Monitoring, .SHIFT = BRIF Shifting, .ORG = BRIEF Organisation, .PLAN = BRIEF Planning, .WM = BRIEF Working Memory, .EMO = BRIEF Emotional Control; AWMA = Automated Working Memory Assessment, DKEFS = Delis-Kaplan Executive Function Battery, TEA-Ch2 = Test of Everyday Attention for Children 2; BRIEF = Behaviour Rating Inventory of Executive Function.

Figure 2

Weight distributions from the self-organizing map, split by assessment. For each assessment, the map depicts high weights (i.e., relatively stronger performance/ratings) as yellow hexagons and low weights (i.e., below age-expected performance/rating) as red hexagons.



Note. Automated Working Memory Assessment: digit recall, backward digit recall, dot matrix, mr x; Delis-Kaplan Executive Function Battery: trails number-letter sequencing & tower; Test of Everyday Attention for Children 2: vigil & reds, blues, bags and shoes; Behaviour Rating Inventory of Executive Function: inhibition; initiation, monitoring, shifting; organisation of materials, planning, working memory, emotional control. Note that all scores are converted to z-scores where scores of zero are equivalent to age-typical levels, scores of 1 indicate one standard deviation above the age-expected mean, and scores of -1 indicate one standard deviation below the age-expected mean.

Figure 3

Distribution of sexes and different diagnoses across data-driven groups.

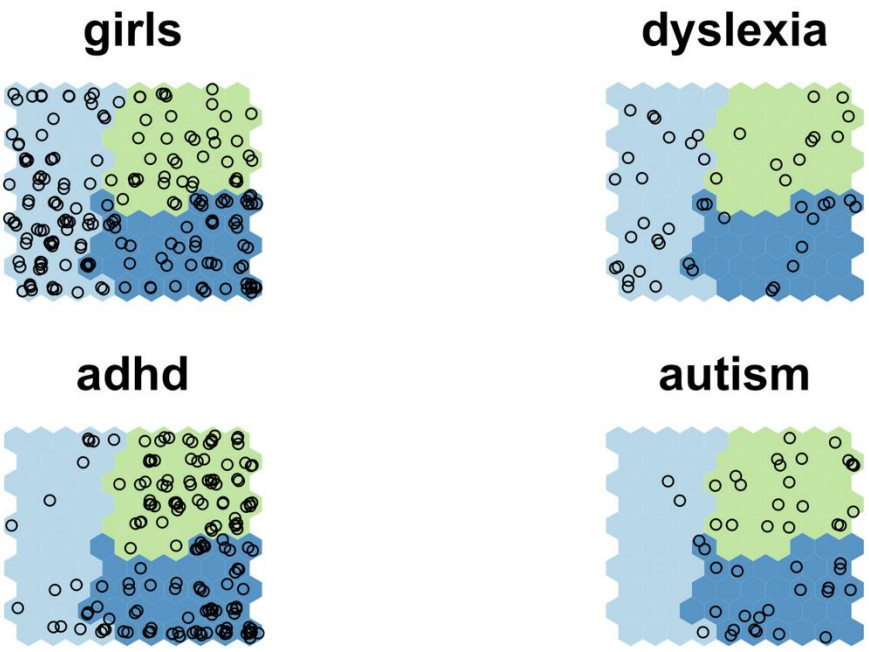
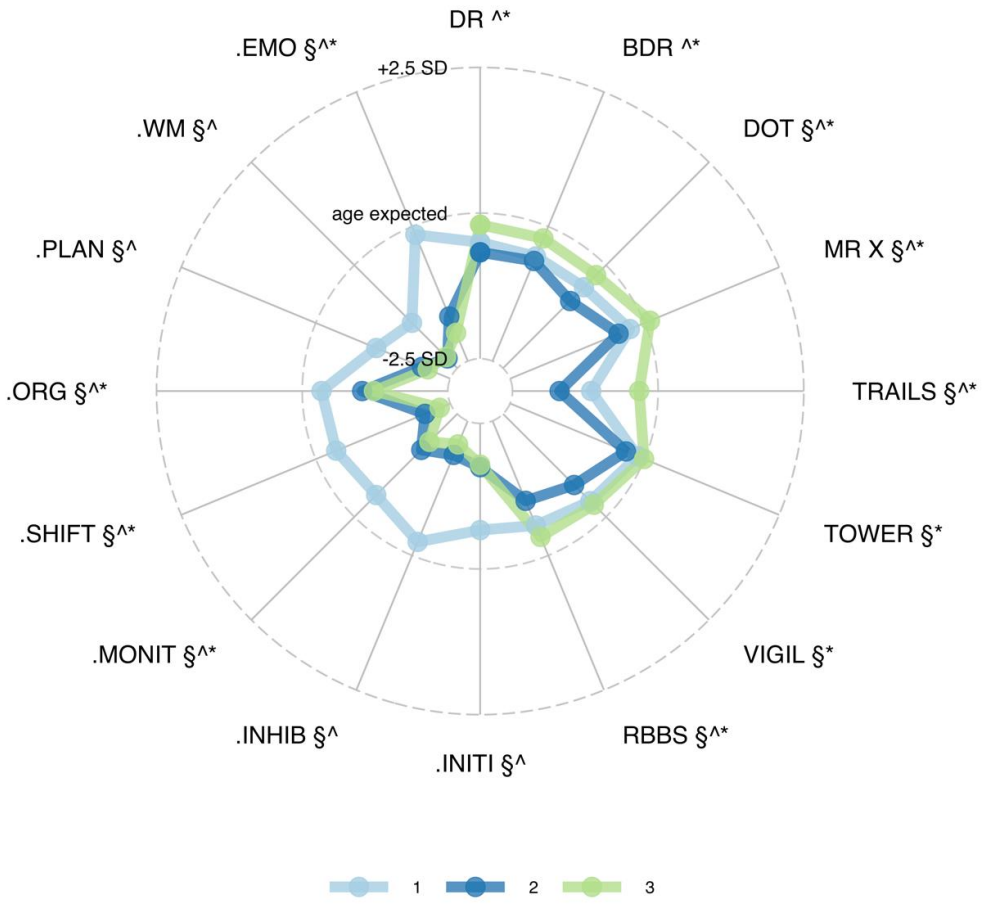


Figure 4

Executive function scores across all assessments used in the self-organising map for each one of the data-driven groups.



Note. All age-referenced scores are converted to z-score (0 = age-expected mean) and p-values are adjusted using false discovery rate correction. DR = AWMA Digit recall, BRD= AWMA Backward digit recall, DOT = AWMA Dot matrix, MR X = AWMA Mr X, TRAILS = DKEFS number-letter sequencing, TOWER = DKEFS Tower, VIGIL = TEA-Ch2 Vigil, SWITCH = TEA-Ch2 Reds, Blues, Bags and Shoes, .INITI = BRIEF Initiation, INHIB =BRIEF Inhibition, .MONIT = BRIEF Monitoring, .SHIFT = BRIF Shifting, .ORG = BRIEF Organisation, .PLAN = BRIEF Planning, .WM = BRIEF Working Memory, .EMO = BRIEF Emotional Control; AWMA = Automated Working Memory Assessment, DKEFS = Delis-Kaplan Executive Function Battery, TEA-Ch2 = Test of Everyday Attention for Children 2; BRIEF = Behaviour Rating Inventory of Executive Function. Note that all scores are converted to z-scores where scores of zero are equivalent to age-typical levels, scores of 1 indicate one standard deviation above the age-expected mean, and scores of -1 indicate one standard deviation below the age-expected mean.

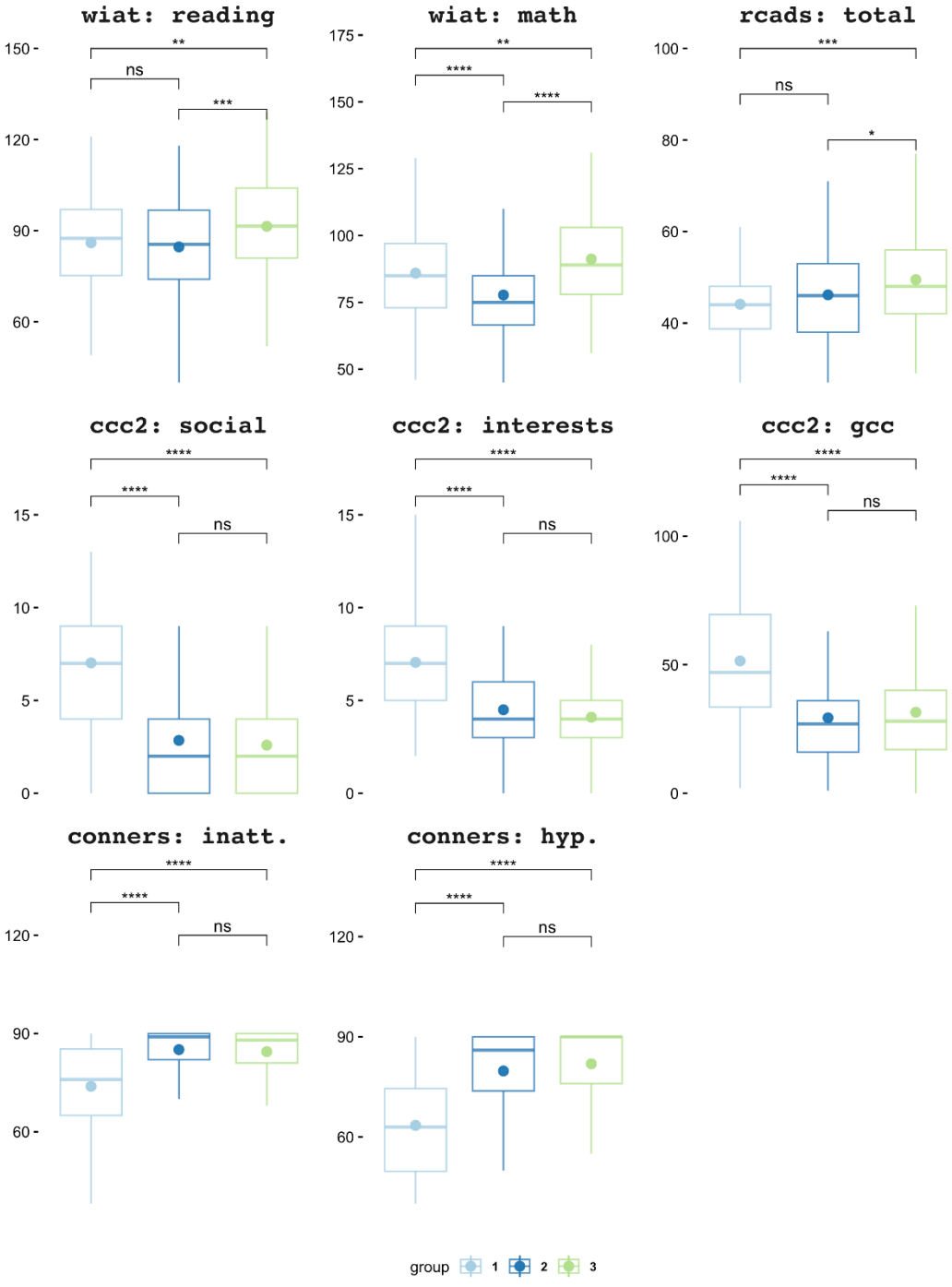
§ = $p_{\text{fdr-corrected}}$ (1 vs 2) <.05

^ = $p_{\text{fdr-corrected}}$ (1 vs 3) <.05

* = $p_{\text{fdr-corrected}}$ (2 vs 3) <.05

Figure 5

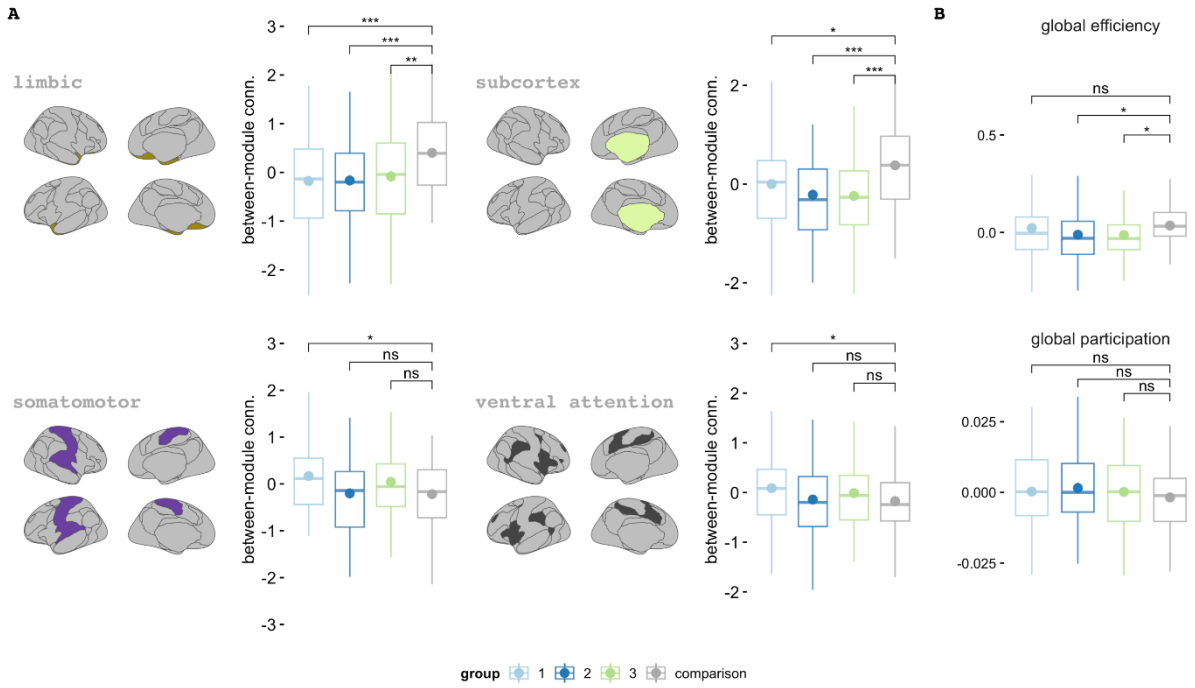
Results of pairwise comparisons of the data-driven groups across measures of learning, self-rated mental health, and patent-rated neurodevelopmental diagnostic dimensions.



Note. WIAT = Wechsler Individual Achievement Test II: reading = Word Reading & math = numerical operations, higher scores indicate relatively better performance; RCADS = The Revised Child Anxiety and Depression Scale (Child-version) total difficulties score, higher score indicate more self-reported difficulties; CCC2 = Children Communication Checklist-2: gcc= general communication composite, social = social relations, higher scores indicate relatively better skills; Conners = Conners-3: inattention and hyperactivity/impulsivity, higher scores indicate relatively more difficulties; All p-values are adjusted using false discovery rate correction.*p<.05, **p <.01, *** p<.001, **** p<.0001

Figure 6

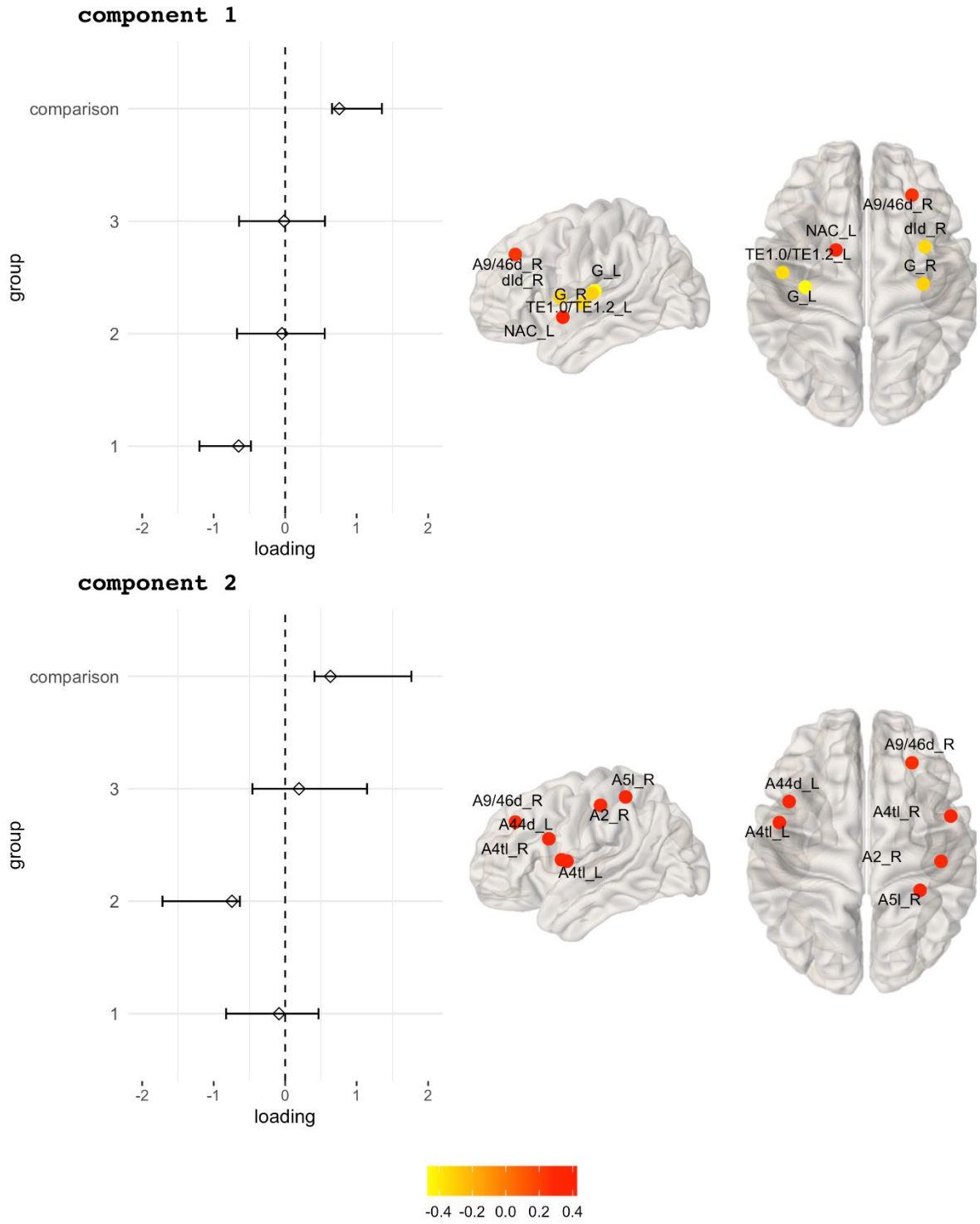
Group differences in modular and global organisation between each one of the three data-driven groups and the comparison group



Note. Panel A shows modules which showed different between-module strength from the comparison sample in at least one of the data-driven groups. Panel B shows group comparisons of global connectome metrics efficiency and participation coefficient. All metrics are age and motion-corrected. All p-values are corrected using false-discovery rate adjustment. *<.05, **<.01, ***<.001

Figure 7

Bootstrapped component loadings of two components identified in the partial least squares discriminate analysis

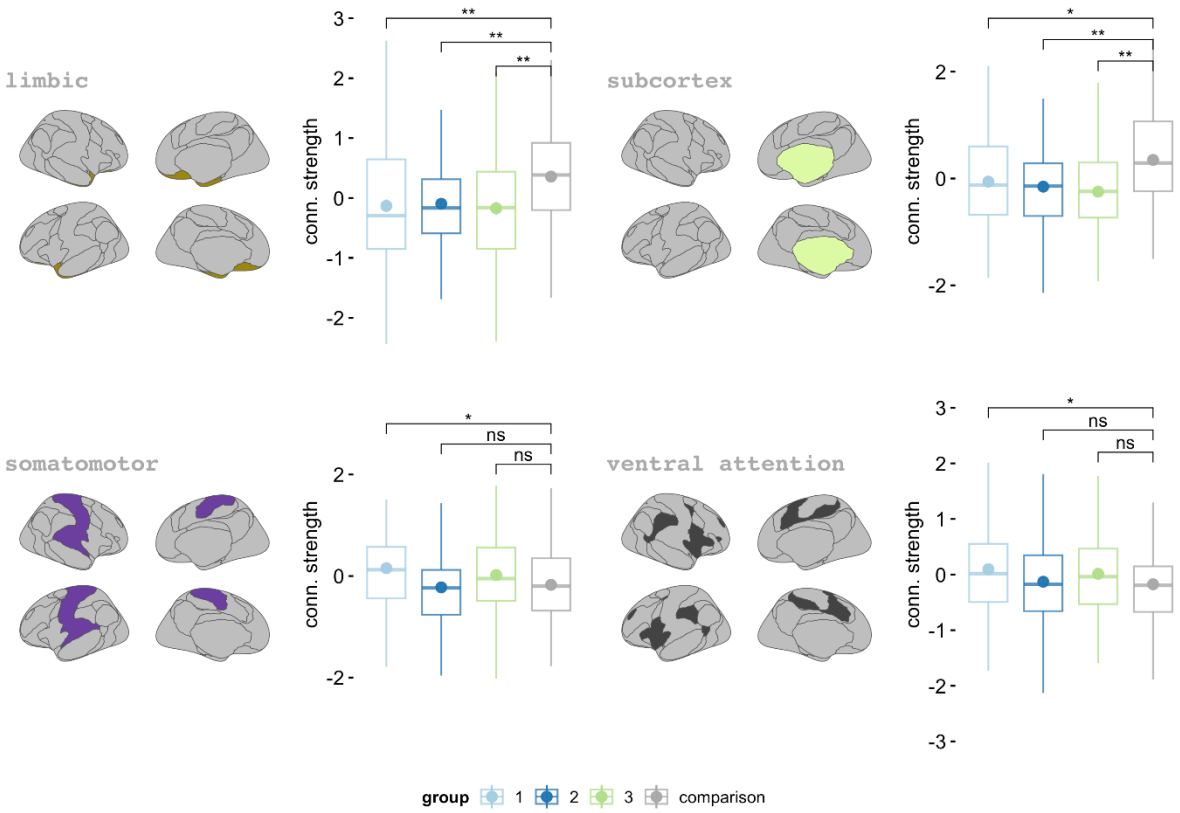


Note. The forest plots show the group loadings for both components (95% confidence intervals based on 1000 bootstraps). Displayed on the left side are connector hub loadings on both components. Only hub loadings that had 95% confidence intervals that did not cross zero are displayed.

Supplementary materials

Figure S1

Group differences in modular strength between each one of the three data-driven groups and the comparison group



Note. The figure displays modules which showed differences in modular connection strength from the comparison sample in at least one of the data-driven groups. . All metrics are age and motion-corrected. All p-values are corrected using false-discovery rate adjustment. *<.05, **<.01, ***<.001