A two-year follow-up study of executive functions in children with Developmental Coordination Disorder

Supplementary Materials

Recruitment procedures and participants. Participants with a diagnosis of Developmental Coordination Disorder (DCD) were recruited for the original study[1] through an advert placed with a charitable organisation, requesting children aged 7-11 with a diagnosis of DCD/dyspraxia to participate in research. Parents volunteered for the study by emailing the research team to receive more information, and to check eligibility. Children with a co-occurring diagnosis of autism spectrum disorder or attention deficit-hyperactivity disorder were excluded from participating due to the potential problems in executive functioning associated with these disorders. Reading and language difficulties, as well as intellectual disability, were assessed through standardised tests (see Materials), and any child demonstrating performance outside of the typical range on these measures was also excluded. The DCD diagnosis was corroborated by the research team using standardised measures and parent report (see Materials). The DCD group in the original study[1] consisted of 23 children (16 males; mean age: 10.0 years, SD: 1.1 years, range: 8.1–11.9). Of these 23 children, 19 agreed to participate in the follow-up study and were re-assessed to ensure that they continued to meet inclusion criteria for the DCD group, and that their diagnosis was stable across time points. Two children scored more than two standard deviations below the mean on the test of intellectual ability. These two children were excluded from the sample (see Table S1 for inclusion/exclusion criteria), because one of the criteria for a DCD diagnosis is that motor deficits are not better explained by intellectual disability (hence the diagnosis could not be corroborated), and because low intellectual ability was likely to impact on their ability to understand task instructions and rules. The final DCD group for the follow-up study consisted of 17 children (11 males; mean age at Time 2: 12.0 years, SD: 1.2 years, range: 10.1 – 14.1).

Children without a diagnosis of DCD were recruited through local schools: parents of 250 children aged 7-11 received information sheets about the study, and volunteered to take part by returning a signed consent form to the research team through the class teacher. Children who did not have any reported medical condition or neurodevelopmental disorder were assessed on the standardised assessments to ensure they met inclusion criteria (see Table S1). Children were included in the typically developing control (TD) group if they scored at or above the 25th percentile on the standardised motor assessment, had no parent-reported motor difficulties, and scored in the typical range on the standardised measures of reading, language and intellectual abilities. Children were identified as having motor difficulties (MD) if they scored at or below the 16th percentile on the standardised motor assessment, but scored in the typical range on the other standardised measures. The original sample[1] included 38 children in the TD group (17 males; mean age: 9.3 years, SD: 1.0 years, range: 7.2–11.1), and 30 children in the MD group (17 males, mean age: 8.9 years, SD: 1.2 years, range: 7.1–11.3). Of these 68 children, 37 were available for follow-up and were re-assessed to ensure they continued to meet inclusion criteria for their assigned group. One TD child performed on the 16th percentile of the MABC-2 and two TD children performed on the 9th percentile. As these children demonstrated some degree of motor difficulty at Time 2 they could no longer be included in the TD group and were therefore excluded from the sample. All children in the MD group continued to meet criteria for group membership, demonstrating persistent motor difficulties across the two time points. The final TD group consisted of 17 children (6 males; mean age at Time 2: 11.3 years, SD: 1.0 years, range: 9.7 – 13.1). The final MD group consisted of 17 children (8 males; mean age at Time 2: 10.5 years, SD: 0.6 years, range: 9.8 – 12.3).

---*Table S1 here*---

Materials. As outlined above, participants were assessed on several standardised measures to confirm their eligibility for the study. These tests are described first, followed by the executive functioning battery.

Movement Assessment Battery for Children (MABC-2) and Checklist. The MABC-2[2] is a standardised assessment of motor ability, comprising three components: manual dexterity, aiming and catching, and balance. Scores for each component can be summed to provide a total standard score (*M*=10, *SD*=3) and percentile ranks, based on UK norms. Children performing at or below the 16th percentile can be identified as having some motor difficulties. Test-retest reliability is reported as .80 for the total sum of the three component scores[2].

The MABC-2 Checklist[2] consists of 30 statements requiring parents to judge their child's level of motor competence in tasks involving movement in a static and/or predictable environment and in a dynamic and/or unpredictable environment, in comparison to other children of the same age. The Checklist is used to assess the impact of motor difficulties on daily life[3], which is central to the diagnostic criteria for DCD. Parents respond to the statements deciding how their child deals with the tasks on a scale from "Very well" to "Not close" (scoring 0–3 points), and a Total Score is calculated. These rating are summed to calculate a total score, which is mapped on three percentile bands, with scores below the 15th percentile representing a risk of motor difficulties and scores below the 5th percentile being indicative of motor difficulties affecting daily living. Test-retest reliability ranged between .77 to .91 in studies using the previous edition of the M-ABC[4], the content of which is highly overlapping with the more recent version.

British Abilities Scales (BAS-3). The BAS-3[5] is a standardised measure of intellectual abilities, comprising both verbal and nonverbal subtests. It was used to ensure that all children were functioning at an appropriate level in order to understand the instructions of the tasks, and to confirm that those in the DCD group did not have an intellectual disability. The Verbal Similarities and Word Definitions subtests were used to measure verbal reasoning, with the Matrices subtest used as a measure of nonverbal reasoning. Scores for each subtest were summed and converted to standard (*T*) scores, with the Matrices *T*-score first doubled to ensure that verbal and nonverbal abilities were equally weighted in the final score (as outlined in the BAS-3 manual). The average of the *T*-scores from the verbal subtests and the doubled nonverbal subtest was calculated and converted into a standard score (General Conceptual Ability [GCA]; M=100, SD=15). Children in all three groups were required to have a GCA score within two standard deviations of the mean (i.e., at or above 70) at both time points in order to be included in the study. Test-retest reliability is reported as .73 for the Matrices subtest, as .86 for the Word Definition subtest and .79 for the Verbal Similarities subtest[4].

Clinical Evaluation of Language Fundamentals 4th Edition (CELF-4-UK). The CELF-4-UK[6], a widely used assessment of receptive and expressive language abilities, was administered to ensure that children did not perform poorly on the verbal executive function measures due to problems with language skills[7], and to exclude children with very low scores indicative of language disorder. Those with scaled scores at or below two SD from the mean (of four or less; M=10, SD=3) on two core subtests, Formulated Sentences (expressive language), and Word Classes-Receptive (receptive language), were excluded. This ensured that children with clear evidence of language disorder did not take part in the study, and that the cut-off harmonised with that used for other study tests (i.e., 2 SD from the mean). Test-retest reliability for relevant ages ranged from .74 to .79 for the Formulated Sentences subtest, and from .83 to .91 for the Word Classes-Receptive subtest[6].

Test of Word Reading Efficiency (TOWRE). The TOWRE[8] was used to assess reading of words and non-words, to ensure that children did not have any reading problems indicative of dyslexia, a disorder that may affect performance on executive functioning tasks[9]. Children were timed when reading a list of words, followed by a list of non-words, and the total number of words read correctly within the time limit of 45 seconds was calculated. Total scores were converted to a standard score (*M*=100, *SD*=15). Children in all three groups were required to have a Total Standard Score within two standard deviations of the mean (i.e. above 70) in order to be included in the study. Test-retest reliability ranged from .82 to .97 for 6 to 9 year-old children[8].

Executive functioning battery. A verbal and a nonverbal test was completed for each of the following executive functions: executive-loaded working memory (ELWM); fluency; inhibition; planning; and cognitive flexibility / switching. A summary of the tasks is provided in Table 2 within the current paper.

For *verbal ELWM*, the Listening Recall test from the Working Memory Battery for Children[10] was completed. Sentences were presented to participants in blocks of six trials, beginning with a block of one-sentence trials, with an increasing number of sentences per trial in each subsequent block. Participants were asked to judge whether the sentence was true or false, and then to hold the last word in memory while providing judgements on the next sentences in the trial. At the end of each trial, children were asked to recall the last words of each sentence in order. The test was ended when three out of six trials within a block were incorrect. Total number of trials correct was scored rather than span, as this has been reported to be a more reliable measure of verbal working memory[11]. Test–retest reliabilities of .38–.83 are reported for relevant ages[10]. For *nonverbal ELWM*, an equivalent test of visuospatial

ELWM was adopted from previous research, called the 'Odd-One-Out' test[12]. On each trial, the child was presented with a card depicting a set of three simple nonsense diagrams and asked to point to the 'odd one out'. Participants were asked to hold the spatial location of the odd-one-out in memory while they provided judgements on the next set of diagrams in the trial. Sets of diagrams were presented in blocks of three, beginning with a block of one-set trials, with an increasing number of sets per trial in each subsequent block. At the end of each trial, children were asked to recall the spatial location of the odd-one-out for each card by pointing to the relevant location on a blank grid. The test was ended when two out of three trials within a block were incorrect. Total number of trials correct was scored. The span version of this task has a reliability of .80[12].

To assess fluency, the Delis-Kaplan Executive Functioning System (D-KEFS[13]) was used. For *verbal fluency,* children were required to generate as many words as possible within one minute that belonged to a specific category (i.e., animals and boys' names). Total correct words (without repetitions) summed from the two categories was used as the measure of verbal fluency. Test-retest reliability is reported as .70 for category fluency[13]. For *nonverbal fluency* ('Design Fluency'), children were provided with a grid in which there were either a number of filled dots (condition one), or a mixture of filled and empty dots (condition two), presented in each square of the grid. Children were required to use four connected straight lines to draw as many different designs as possible within one minute. In condition two, children were only allowed to connect the empty dots. Nonverbal fluency was calculated using the total correct designs (i.e. those following the rules) across the two conditions. Test-retest reliabilities are reported as .66 for filled dots and .43 for empty dots[13].

To assess inhibition, a test was adopted from previous research[7] called the Verbal Inhibition, Motor Inhibition (VIMI) test. For *verbal inhibition*, children were required to repeat words said by the experimenter (i.e., either 'doll' or 'car'), which were presented in a pseudo-random order for 20 trials (‘copy’ block). For the next block of 20 trials (‘inhibit block’), participants were required to inhibit this copying response by responding with the opposite word (i.e., ‘car’ was the response to ‘doll’, and vice versa). The copy and inhibit blocks were then repeated once with the same words (Part A), followed by a set of four blocks following the same pattern but using different words (‘bus’ and ‘drum’; Part B). Total number of errors across the full task provided the measure of verbal inhibition. Cronbach’s alpha, based on total error scores, was .73[7]. For *nonverbal inhibition*, the test followed an identical format but used hand actions instead of words. Participants were required to copy the experimenter in presenting a pointed finger or a fist (Part A), or a flat horizontal hand or flat vertical hand (Part B). In the ‘inhibit’ blocks, participants again had to present the opposite hand action to the experimenter. Total number of errors across the full task provided the measure of nonverbal inhibition, and Cronbach’s alpha for these error scores was .92[7].

To assess planning, the D-KEFS Sorting task[13] was used. Participants were presented with two sets of six cards and asked to sort them into two groups of three in as many different ways as they could. Categories could be created based on the words presented on the cards (*verbal planning*), or on the perceptual properties of the cards (*nonverbal planning*). There were three possible verbal sorts (e.g., transports *vs.* animals, things that fly *vs*. things that move along the ground) and five possible nonverbal sorts (e.g., small cards *vs.* large cards, straight edges *vs.* curved edges) in each card set. Total numbers of correct sorts were used as the measures of verbal and nonverbal planning, respectively. Test-retest reliability for the Sorting task is reported as .49[13].

To assess cognitive flexibility, two tasks were adopted from standardised batteries of executive functioning measures. For *verbal cognitive flexibility*, the D-KEFS Trail Making Test[13] was used. In the number-letter switching task, participants were required to connect letters and numbers in an alternating sequence (i.e., 1-A-2-B-3-C, etc., until 16-P) as quickly as possible. In order to ensure that reduced performance on this task was not caused by difficulties with sequencing numbers or letters, or due to motor speed or visual scanning abilities, component skills were also assessed. In the motor speed task, children were required to follow a line with their pencil between dots placed around the page (as in a 'dot-to-dot' game) as quickly as they could, thus removing any of the verbal element from the task. In the visual scanning task, children were asked to find all the number 3s on the page and cross them off as quickly as possible. The number sequencing task involved connecting the numbers from 1-16, and the letter sequencing task required connecting the letters from A-P. The measure of verbal cognitive flexibility was the total time for the number-letter switching task minus the total time for the number and letter sequencing tasks (i.e., 'switching cost'). Test-retest reliabilities for the component tasks are reported as .77 (number sequencing), .57 (letter sequencing) and .22 (letter-number switching)[13]. The fact that switching measures depend on difference scores can make reliability of these tasks somewhat low, but this is an inherent problem with these measures[14]. For *nonverbal cognitive flexibility*, the Intra-Extra Dimensional Set Shift test from the Cambridge Neuropsychological Test Automated Battery (CANTAB)[15] was used. Participants were first presented with two coloured shapes and asked to work out the rule by touching one of the two shapes on the screen and finding out whether they were 'correct'. Feedback was provided by the computer program, and participants were told that if they had found the correct shape, they should continue to touch this shape on subsequent trials until the rule changed (i.e., until they received feedback that their response was 'incorrect'). At this point children would need to switch rule, and choose the other shape instead. In the second part of the task, a white line was added to the stimuli, either adjacent to or overlaying the coloured shape, but the child continued to attend to the coloured shape to obtain correct responses ('intra-dimensional shift'). In the final part of the task, the rule changed again and children had to attend to the white line in order to obtain correct responses (‘extra-dimensional shift’), ignoring the coloured shape to which they had previously been attending. Total number of errors across the task was used as the measure of nonverbal cognitive flexibility. Test-retest reliability for total errors is reported as .40[15].

Statistical Analyses. Statistical checks in each regression (e.g. Durbin-Watson, variance inflation factor statistics, standardised residuals, Cook’s/Mahalanobis distances) revealed no evidence of multicollinearity and no outliers or influential cases[16].

Supplementary References

[1]Leonard HC, Bernardi M, Hill EL, Henry LA. Executive functioning, motor difficulties, and developmental coordination disorder. *Dev Neuropsychol* 2015; 40(4):201-15.

[2]Henderson SE, Sugden DA, Barnett AL. Movement assessment battery for children (Movement ABC-2) (2nd ed.). London: The Psychological Corporation, 2007.

[3]Schoemaker MM, Niemeijer AS, Flapper BC, & Smits-Engelsman BCM. Validity and reliability of the Movement Assessment Battery for Children-2 Checklist for children with and without motor impairments. *Developmental Medicine & Child Neurology* 2012; 54(4): 368-375.

[4]Henderson SE, Sugden, DA. The Movement Assessment Battery for Children (Movement ABC). London: The Psychological Corporation, 1992.

[5]Elliot CD, Smith P. British Ability Scales – Third Edition (BAS-3). London: GL Assessment, 2011.

[6]Semel E Wiig EH, Secord WA. Clinical Evaluation of Language Fundamentals CELF-4 UK (4th ed.). London, UK: Psychological Corporation, 2006.

[7]Henry LA, Messer DJ, Nash G. Executive functioning in children with specific language impairment. *J Child Psychol Psychiatry* 2012; 53(1):37-45.

[8]Torgensen J, Wagner R, Rashotte C. Test of Word Reading Efficiency (TOWRE). Austin, TX: Pro-Ed, 1999.

[9]Booth JN, Boyle JME, Kelly, SW. Do tasks make a difference? Accounting for heterogeneity of performance of children with reading difficulties on tasks of executive function: Findings from a meta-analysis. *Brit J Dev Psychol* 2010*, 28,* 133-176.

[10]Pickering S, Gathercole SE. Working memory test battery for children (WMTB-C). London: The Psychological Corporation, 2001.

[11]Ferguson AN, Bowey JA, Tilley A. The association between auditory memory span and speech rate in children from kindergarten to sixth grade. *J Exp Child Psychol*, 2002; 81: 141–156.

[12]Henry LA. How does the severity of a learning disability affect working memory performance? *Memory* 2001; 9(4-6):233-47.

[13]Delis DC, Kaplan E, Kramer JH. Delis–Kaplan Executive Function System (D-KEFS®). London: Psychological Corporation, 2001.

[14]Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex ‘‘frontal lobe’’ tasks: A latent variable analysis. *Cogn Psychol*, 2000; 41: 49–100.

[15]Cambridge Cognition. The Cambridge Neuropsychological Test Automated Battery (CANTAB®): Test–retest reliability characteristics. Cambridge: Author, 2006

[16]Field AS. Discovering Statistics using IBM SPSS Statistics (4th ed.). London: Sage Publications Ltd, 2013.

Table S1

Inclusion criteria for group membership at Time 1 and Time 2

Inclusion Measure	TD group	MD group	DCD group
Movement Assessment Battery for Children (MABC-2) and Checklist	MABC-2 Total score ≥ 25th %, Checklist > 15th %	MABC-2 Total score ≤ 16th %	MABC-2 Total score ≤ 16th %, Checklist < 5th %
British Abilities Scales (BAS-3)	Standard score ≥ 70	Standard score ≥ 70	Standard score ≥ 70
Clinical Evaluation of Language Fundamental (CELF-4-UK)	Scaled score ≥ 4 on Formulated Sentences and Word Classes-Receptive subtests	Scaled score ≥ 4 on Formulated Sentences and Word Classes-Receptive subtests	Scaled score ≥ 4 on Formulated Sentences and Word Classes-Receptive subtests
Test of Word Reading Efficiency (TOWRE)	Standard score ≥ 70	Standard score ≥ 70	Standard score ≥ 70
Parent reports of clinical diagnosis	No clinical diagnosis	No clinical diagnosis	Diagnosis of DCD only