The Assessment of Executive Functioning in Children

Lucy A. Henry & Caroline Bettenay

Child and Adolescent Mental Health, 2010, 15(2), pp. 110-119.

Department of Psychology, London South Bank University, 103 Borough Road, London SE1 0AA, UK. E-mail: henrylc@lsbu.ac.uk

Background: Executive functioning is increasingly seen as incorporating several component sub-skills and clinical assessments should reflect this complexity. **Method:** Tools for assessing executive functioning in children are reviewed within five key areas, across verbal and visuospatial abilities, with emphasis on batteries of tests. **Results:** There are many appropriate tests for children, although the choice is more limited for those under the age of 8 years. **Conclusions:** Whilst there are several batteries of executive functioning suitable for children, clinicians may prefer to cherry-pick from a broader range of measures that assess specific components of executive functioning.

Keywords: Executive functioning; children; assessment

Background

Executive functions are 'those skills necessary for purposeful, goal-directed activity' (Anderson, 1998), required for the successful achievement of complex, higher order cognitive goals, including planning future actions, keeping these plans in mind until executed, problem-solving, self-monitoring to check on progress, mental flexibility, and the ability to inhibit irrelevant actions.

Executive control is a prominent part of the influential multicomponent model of working memory (Baddeley & Hitch, 1974; Baddeley, 1986, 2000, 2007) and involves focusing, dividing and switching attention. Executive abilities are increasingly regarded as at least partly 'fractionated' or divided (e.g. Lehto, 1996) into separate subcomponents that are, nevertheless, loosely related to each other (Anderson, 2002; Fisk & Sharp, 2004; Garon, Bryson, & Smith, 2008; Huizinga, Dolan, & van der Molen, 2006; Lehto et al., 2003; Miyake et al., 2000). Executive control is not used during routine (automatised) tasks (Shallice, 1990), but for demanding tasks that involve novelty.

Pennington and Ozonoff (1996) divided executive functioning into discrete sub-skills that included planning, working memory/updating, problem-solving, self-monitoring, mental flexibility, generativity/fluency, and inhibition of prepotent responses. Miyake et al. (2000) argued that three key aspects of executive functioning in adults (inhibition, updating and shifting) were 'separable but moderately correlated constructs' (p. 87). Further studies have broadly supported these conclusions in adults and children (Anderson, 2002; Fisk & Sharp, 2004; Garon et al., 2008; Lehto et al., 2003), although there are still uncertainties regarding the precise nature and specification of executive abilities.

Executive functions develop slowly from infancy (e.g. Diamond & Goldman-Rakic, 1989) through early childhood and adolescence, and may still improve into young adulthood (e.g. Anderson, 2002; Garon et al., 2008; Levin et al., 1991; Huizinga et al., 2006; Welsh, Pennington, & Groisser, 1991). Executive functioning is a key cognitive skill underpinning successful goal-directed behaviour, and is linked to educational attainment in English, maths and science (St. Clair-Thompson & Gathercole, 2006). Executive dysfunction refers to deficits in the ability to inhibit well-learned patterns of behaviour and derive new ways of solving problems. Individuals become trapped in repetitive cycles of well-learned behaviour (perseveration) and lack flexibility to accommodate and re-accommodate their behaviour to novel situations.

Measuring executive functioning in children

This review will focus on five standardised batteries of executive functioning (see Table 1 at the end of the paper for a summary) designed for school-age children (for information on executive functions in preschoolers, see Garon et al., 2008). Verbal *and* visuospatial domains are distinguished to allow for domain specific comparisons. Measures of executive functioning are considered in five areas: (1) *Executive-loaded working memory*. Working memory is 'a system for temporarily holding and manipulating information as part of a wide range of essential cognitive tasks such as learning, reasoning and comprehending' (Baddeley, 1997). The key feature in assessing executive-loaded working memory is requiring *both* processing and storage of the results of that processing, often measured using 'complex span' tasks. (ii) *Fluency/reconstitution*. These measures require participants to generate items around a particular theme (e.g. verbal concepts, ideas or visuospatial criteria), to test the efficiency and flexibility of search processes. (iii) *Inhibition*. This refers to 'the deliberate, controlled suppression of prepotent responses' (Miyake et al., 2000). (iv) *Set shifting/switching*. These measures require the ability to change/adapt mental set when required, including the ability to change/alternate a strategy in a responsive manner, or abandon a strategy in response to negative feedback. (v) *Planning/problem-solving*. This emphasises the person's ability to develop goals, work out strategies, monitor performance and generate new solutions.

Why measure executive functioning in children? There is mounting evidence that many children with developmental disorders have particular *profiles* of executive impairment, which may help practitioners develop treatment programmes. For example, planning and set shifting are impaired in autism (Hill, 2004; Liss et al., 2001), whereas inhibition appears to be a primary deficit in attention deficit hyperactivity disorder (Barkley, 1997). Although patterns of impairment are not yet well specified in many other disorders such as Down syndrome, Williams syndrome, and specific language impairment, executive skills in children with a range of developmental disorders are receiving increasing attention (e.g. Bishop & Norbury, 2005a; 2005b; Channon, Pratt, & Robertson, 2003; Ellis-Weismer, Evans, & Hesketh, 1999; Guerts et al., 2004; Hill, 2004; Lanfranchi, Cornoldi, & Vianello, 2004; Liss et al., 2001; Ozonoff & Jensen, 1999; Pennington & Ozonoff, 1996; Robinson et al., 2009). Given that executive skills are central to effective goal-oriented behaviour, they are worth assessing in children who have cognitive and/or behavioural difficulties.

This review examines several test batteries, plus some additional measures of executive functioning, to give practitioners an overview of the range of tests available, as well as the executive functioning subcomponents they measure. A summary table at the end of this article allows comparisons between batteries.

Batteries of executive functioning

The Delis-Kaplan Executive Function System (D-KEFS, Delis, Kaplan, & Kramer, 2001). The D-KEFS is a comprehensive test of executive functioning in children from 8 years (Proverbs Test suitable only for adults), which is reasonably straightforward to administer. It includes good controls for component skills in several tasks, good information on reliability and validity; and many instances of separate scores for verbal and perceptual domains. Of the nine subtests, seven are based on traditional executive functioning measures, which the authors argue add to its validity. Many measures are clearly divisible into verbal and visual spatial domains, either because there are parallel analogous tests, or because these aspects are measured separately. This is useful for evaluating individuals in whom there are suspected differences between verbal and visuospatial abilities. The D-KEFS provides age-scaled scores and detailed analyses of errors and contrast scores.

Standardisation involved 1750 nationally representative American children (75 to 100 at each age level 8-15 years) and adults. Information on reliability and validity is thorough; with adequate to good test-retest reliabilities and internal consistency coefficients for most measures (although the complexity of executive function measures may increase performance variability and measurement error, such that high values cannot be expected, Delis et al., 2004), with some exceptions (some switching measures). Alternate forms for three measures are available (Sorting, Verbal Fluency, 20 Questions) and there are guidelines in the manual as to clinical interpretation of D-KEFS subtests (Homack, Lee, & Riccio, 2005).

Verbal domain

Verbal fluency (e.g. Thurstone Word Fluency Test, Thurstone, 1938; Milner, 1964) in the D-KEFS includes letter, category (e.g. generate animal names) *and* switching (between categories) measures of fluency; they are quick and straightforward to administer with modest to good internal consistency (.37-.80) and slightly better test-retest reliability (.53-.70). The complex scoring provides plenty of clinical detail. Note that category and letter fluency have been linked to different brain areas (Martin et al., 1994; Schwartz et al., 2003), implying that different mechanisms may underlie performance.

Inhibition is often measured using the Stroop Test (Stroop, 1935), requiring the ability to inhibit a habitual response. Colour words ('red', 'blue') are presented in different coloured inks and the participant names the colour of the INK rather than name the word. The 'habitual' word-naming response must be inhibited, hence this test may be considered verbal. The D-KEFS version has good control conditions to vary out relevant component skills (colour naming speed, word reading speed), so would be suitable for individuals in whom speed of word naming and reading may be compromised. Internal consistency is moderate to reasonably high (.62-.79), with good to high test-retest reliabilities for children and adolescents (.77-.90).

The Trail Making Test (Reitan & Wolfson, 1992), a measure of set shifting/switching, also assesses aspects of speed of visual search, attention, and visuo-motor function. Part B requires participants to draw lines between letters and numbers alternately in sequence (for example A to 1; 1 to B; B to 2; 2 to C etc) and is, therefore, a test of switching. The D-KEFS version includes impressive control conditions (speed of visual scanning, number sequencing; letter sequencing; motor speed), making it suitable for individuals who have motor difficulties. Moderate to good internal consistency (.57-.79) contrasts with rather variable test-retest reliability (.20-.82) across the five conditions but, unfortunately, the lowest figure represents the measure of greatest interest, switching. The task is easy to administer, and although scoring is complex, it provides a useful range of detailed scores.

The Sorting Test from the D-KEFS is one of the few planning tasks available that includes a verbal element, yielding measures of organisational skill in verbal and visuospatial domains. Participants sort a set of six items that look like puzzle pieces into two piles in as many ways as possible (e.g. colour, shape, background design, category), with five possible perceptual sorting rules and three possible verbal sorting rules. Participants sort spontaneously; they are then shown sorts they must describe. Internal consistency values (.55-.80) and test-retest reliabilities (.49-.67) demonstrate moderate to relatively good reliability. Like many D-KEFS subtests, scoring is complicated, but delivers a rich array of detail. There are separate verbal and perceptual scores for correct *sorts* and correct *descriptions*. One weakness with this test is that children in the younger age range find it quite difficult.

Visuospatial domain

Jones-Gotman and Milner (1977) developed a Design Fluency Test, requiring participants to invent different 'nonsense' drawings; the D-KEFS includes an adapted version of this test. Sets of identically placed dots must be connected, differently each time using four-lines, to produce different diagrams (like the same mini 'dot to dot' puzzle, without the numbers, completed over and over again). This task is more constrained than previous measures, because the sets of dots provide a structure for drawings; a switching condition is also included (between filled and empty dots). Internal consistency data are not available, but test-retest reliability was moderate in the child sample (.43-.66). Together with verbal fluency, the D-KEFS provides assessments of fluency/switching within both verbal and visuospatial domains.

Ideational fluency measures require participants to think of as many uses for an everyday object as possible (e.g. a brick or a newspaper), or think of interpretations for meaningless patterns (Turner, 1999; Bishop & Norbury, 2005a). Anderson (1998) describes the Twenty Questions test (20 'yes-no' questions to identify a target) as a measure of concept formation. One such standardised measure is available in the D-KEFS, but the test-retest reliability value for 'overall achievement' on the test is virtually zero (.06) so caution should be exercised in interpreting this test (although the 'initial abstraction score' test-retest reliability coefficient was higher, .62).

Tower tests (e.g. Tower of London, Tower of Hanoi) are commonly used measures of problem-solving, developed to minimise the contributions of perceptual and motor abilities, short-term memory and sustained attention. They require the rearrangement of coloured balls (or different sized discs) from an initial starting point on three laterally placed 'posts' or 'stockings' to a specified end point in the minimum number of moves. Some suggest these tasks measure *more* than planning as they require inhibition of 'obvious' or impulsive moves that bring the participant superficially closer to the goal, but are unhelpful for the longer-term solution. Miyake et al. (2000) argue that if participants use a demanding 'goal management' strategy involving setting up subgoals, maintaining them in short-term memory and executing them sequentially, this task is a measure of planning. However, if participants use a simpler 'perceptual' strategy, making successive moves that lead to the display 'looking' more like the desired end state, this task is more a measure of inhibition. Therefore, clinicians using tower tasks should allow for strategic differences when interpreting scores. A standardised Tower Test is available in the D-KEFS, with modest to good internal consistency (based on correlating two ‘half’ tests, .43-.84) and moderate test-retest reliability (.51).

In summary, the D-KEFS includes several tests that incorporate better controls for component skills than earlier clinical versions (e.g. Stroop, Trail Making Test). The authors claim that in using a range of ‘long-standing clinical or experimental tests’ (p.57, *Technical manual*, Delis et al., 2001) with a history of over 50 years of neuropsychological research, the D-KEFS has demonstrably good construct validity. Evidence of its sensitivity in distinguishing clinical groups has been presented (Homack et al., 2005). Discriminant validity was demonstrated via correlations between D-KEFS subtests and a verbal memory test battery - these were largely non significant, as would be expected. Further, correlations between the nine sub-tests of the D-KEFS and the Wisconsin Card Sorting Test were modest to moderate (.31-.59). Note that we would not expect correlations to be high between different measures of executive functioning; as these measures are believed to be only loosely related (e.g. Miyake et al., 2000). Two new tests were developed in the D-KEFS battery (Sorting Test, Word Context Test); the authors quote several studies using a version of the Sorting Test (the California Card Sorting Test, Delis, 1988) attesting to its validity.

Homack et al. (2005) described some potential weaknesses with the D-KEFS: it requires close monitoring by the examiner during administration; some test instructions are rather complex and repetitive; although the record form is well-organised, hand scoring is quite laborious; and, as factor-analytic methods were not used to develop the D-KEFS, empirically derived factor scores are not available (although the authors point out serious limitations to this factor analytic approach, Delis et al., 2004).

The Cambridge Neuropsychological Test Automated Battery (CANTAB) (Cambridge Cognition Ltd, 2006a). This computerised assessment covers executive functioning, working memory and planning. Normative data are available from 4 years, so this test is suitable for young children. Of the 23 tests in the CANTAB, most (21) are nonverbal both in terms of test presentation and participant response, so it is valuable for participants with limited verbal abilities and may be more 'culture-free' than other tests. Test presentation is via a touch screen computer, which may be advantageous for certain clinical groups (e.g. those with autism, Luciana, 2003) and all responses are recorded and analysed automatically. The testing format is suitable for children and adults if such comparisons are needed, and administration can be carried out by a trained assistant.

Construct and discriminant validity of the CANTAB in child populations has been demonstrated in a range of research studies using clinical and non-clinical samples for over 15 years (e.g. Curtis, Lindeke, & Georgieff, 2002; Hughes, Plumet, & Leboyer, 1999; Luciana, 2003; Luciana & Nelson, 1998), although questions have arisen as to the ability of the Stockings test (similar to Tower tests described above) to discriminate levels of ability in high functioning individuals with autism over the age of 12 (Ozonoff et al., 2004). A reliability study with healthy adult participants (N=100, mean age = 44.1 years), found largely modest to good test-retest reliabilities (.4 - .87) (Cambridge Cognition Ltd, 2006b). Reliability data are not available for children, but internal consistency coefficients were reported as high (.73-.95) in a sample of 4 to 12-year-old children (Luciana, 2003). Four tests (Spatial Working Memory, Stop Signal Test, Intra-Extra Dimensional Shift, Stockings of Cambridge) are suitable to assess executive functioning in the visuospatial domain.

Visuospatial domain

Spatial Working Memory, a measure of executive-loaded working memory, requires participants to retain and use efficient strategies to manipulate spatial information. The child finds tokens hidden among squares. Each square can only be used once and the next hiding place is unpredictable. If the child returns to a square that has already held a target, this is an error. Task difficulty ranges from two-item searches to eight-item searches and scores are computed automatically for strategy efficiency and memory errors, and compared to norms. Test-retest reliability on the adult sample (CANTAB Reliability Study) was satisfactory (.70 for errors, .63 for strategy use); but data are unavailable for children.

The Stop Signal Task assesses inhibition; participants press a button on the left when they see a left-pointing arrow and a button on the right when they see a right-pointing arrow. During inhibition trials, participants continue pressing the buttons in response to arrows, but must withhold their response when they hear an auditory signal. The test is straightforward and easy to administer, but there are no data available on reliability in children or adults, as it is new to the battery (see Williams, et al., 1999, for reference norms).

An excellent example of an easy to administer switching measure that is clearly in the visuospatial domain is the Intra-Extra Dimensional (IED) shift task, one of the only measures of nonverbal switching currently available. Participants choose one of a pair of nonsense line drawings in different colours or broad shape outlines, with feedback on whether the choice is ‘correct’ or not. At certain points, the dimensional criteria are changed without warning and the participant must discover the new rule. At other points, two co-occurring dimensions are introduced. Modest to good reliability values, based on the adult test-retest study, were .40 for total errors and .75 for stages completed. One weakness of this test is that it can be rather lengthy for younger participants.

Finally, Stockings of Cambridge is the CANTAB version of a ‘Tower Test’, suitable for children who prefer computer-based presentation. Test-retest reliability for this test was moderately good (.64), again based on adult data.

Although the CANTAB has many strengths and is attractive for children, Luciana (2003) outlines some potential weaknesses: it is very expensive to purchase; certain populations may perform differently in computerised versus standard assessments (e.g. children with autism may perform better with computerised tests); the purity of the tasks and assessment format may reduce their ecological validity; and the emphasis on nonverbal stimulus presentation/response requirements means that clinicians cannot comment on comparable verbal measures.

The Test of Everyday Attention for Children (TEA-Ch, Manly et al., 1999). This battery was developed primarily as a measure of attention, minimising the contributions of memory, reasoning, task comprehension, motor speed, verbal ability and perceptual acuity (Manly et al., 2001). Age-scaled scores and percentiles are available for boys and girls across 6 age bands (6-16 years). Manly et al. (2001) report normative data on 293 healthy UK children, and reasonable test-retest reliability in a random subgroup of 55 children (range .65-.85). A parallel version is available with separate scoring to accommodate practice effects. Assessment of convergent validity indicates promising relationships between other measures of executive function and

TEA-Ch subtests (see below); evidence for discriminant validity was provided as correlations were not present with IQ; and boys with ADHD performed poorly on the TEA-Ch, as would be expected. Of the nine subtests in the TEA-Ch, several implicate executive functioning (Walk Don't Walk, Creature Counting, Opposite Worlds, Sky Search Dual Task).

Verbal domain

The Opposite Worlds Test measures inhibition. Children see a stimulus sheet with a snaking pattern of digits, like a board game (1s and 2s semi-randomly presented). In the congruent condition (Sameworld) children read out the digit names as quickly as possible. In the incongruent condition (Oppositeworld), children read out the opposite name for each digit as quickly as possible. This task requires inhibition of prepotent 'correct' digit names and this is the dependent measure of interest. Test-retest reliability is excellent (.92); evidence for validity comes from modest correlations with the Stroop test (.24) and the Matching Familiar Figures Test (.25) (Manly et al., 2001).

Visuospatial domain

In the Walk Don't Walk test of inhibition, children move along a path made up of 14 squares by 'dotting' each square with a marker pen. The signal to make a move forward and place a dot in a square is an auditory tone. The signal to *not* make a move forward is an identical tone to start with, but one with a different ending. This requires the child to listen to the *full tone* to decide whether to go forward or not. This task has reasonably good test-retest reliability (.73), as well as being modestly related to two other measures of executive function (Trails B r=.3; Matching Familiar Figures Test, r=.20), providing evidence for its validity (Manly et al., 2001).

Visuospatial and verbal domains

A measure of switching, Creature Counting, involves children counting variable numbers of creatures hiding in snake-shaped burrows. They switch between counting forwards and backwards every time they reach an arrow, interspersed at regular intervals between the creatures. The arrow points up to signal forward counting or down to signal backwards counting. Creature Counting has reasonably good test-retest reliability (.69 accuracy, .73 timing) and evidence for validity is provided by modest but positive correlations with the Stroop task (.31), Trails B (.21) and the Matching Familiar Figures Test (.35) (Manly et al., 2001). By including verbal counting and visuospatial arrow symbols, this task requires both verbal *and* visuospatial processing.

In the Sky Search Dual Task, children carry out a relatively realistic and interesting visual search task (searching for identical pairs of spacecraft on a large laminated sheet containing over 120 pairs of spacecraft, most of which are *not* identical pairs), while silently counting identical auditory tones over 10 trials, and relaying the total counts on each trial. Scores reflect the decrement produced by carrying out the two tasks simultaneously; a measure of dual task interference. Although convergent validity for this task was weak (scores were not related to the Stroop Test or the Trail Making Test), test-retest reliability was good (.81, Manly et al., 2001). Again, this task requires both verbal and visuospatial processing.

TEA-Ch subtests have been used in many research studies, although there have been reports of children with developmental disorders finding some subtests unpalatable (Bishop & Norbury, 2005b, Walk Don't Walk subtest). Other than Creature Counting, which the authors admit is sometimes difficult to explain, the tests are straightforward to administer and score. Not all subtests are scored in the same way and baseline measures of speed and accuracy are not recorded (unlike D-KEFS), which may be problematic in interpretation of scores (Wilding, 2005). Some tasks are not pure measures of either visuospatial or verbal processing.

The Behavioural Assessment of the Dysexecutive Syndrome in Children (BADS-C, Emslie et al., 2003). The key feature of this battery is its claim to measure executive functioning in an ecologically valid manner. It is suitable for children and young people of 8-16 years. Normative data (265 healthy UK children, nationally representative of ability and socio-economic group, plus 114 children with developmental/neurological disorders) were supplemented with a small sample (22) of healthy 7-year-olds in 2006. The battery includes the Dysexecutive Questionnaire for Children (DEX-C), a 20-item measure of the types of problems commonly associated with dysexecutive syndrome. Age-scaled scores and percentiles are provided for each year band.

Visuospatial domain

Planning is measured using several tests (Water, Key, Zoo Map, Six Part). The Water Test is argued to be an 'ecologically valid' test of planning, requiring children to remove a cork from a tube using an array of physical objects and materials (water, plastic tube, screw top, cork); five correct interim steps are needed to succeed. The

Key Test looks at how well children can plan an efficient and systematic search of a 'field' in which they have lost their keys. The 'field' is an A4 piece of paper and the 'search' must begin from a particular point and be marked out by drawing lines with a pen. Thinking ahead in both tasks is essential, hence they should reflect planning. Lack of research makes these tests hard to compare with similar measures, but they are appealing to children and do appear to assess the visuospatial domain. Test-retest reliability is unavailable for the Water Test as all children reached ceiling on the second testing; for the Key Search Test, test-retest reliability was good (.81), albeit on a small sample (25). Information on validity is limited, but the test manual notes low to minimal correlations between the BADS-C subtests and other measures of executive function (e.g. TEA-Ch).

Although there are two further planning tests in the BADS-C (the Zoo Map Test involves planning a route to visit six out of 12 locations in the zoo in accordance with a set of rules and the Six Part Test requires planning, task-scheduling and performance monitoring), both require a complex combination of verbal and visuospatial skills so cannot be classified by processing domain. Set shifting/switching is measured using the Playing Cards Test, which requires the child to respond according to one rule and then change strategy in accordance with a new rule, but its test-retest reliability was extremely poor (-.24).

The ecological validity of the BADS-C is an advantage for clinicians wanting to examine everyday difficulties in executive functioning, although it is difficult to separate verbal and visuospatial skills in many tests. Test-retest reliabilities are highly variable, which the authors argue is a reflection of the need to provide novel tasks assessing EF, but some variability could be accounted for by the small sample size (25). Test performances may be difficult to interpret because there is little research on these tasks to determine what types of executive functions precisely they measure. Most BADS-C measures load more heavily on non-verbal abilities, which may be an advantage when testing certain populations, although if verbal difficulties are suspected, this battery would be best used in combination with other verbal executive measures. Although there were significant correlations between five of the six BADS-C subtests and the DEX-C within a typical sample, correlations between subtests and other executive measures such as the TEA-Ch in clinical sample were largely non-significant or low.

The NEPSY II (Korkman, Kirk & Kemp, 2007). This is a revision and extension of the original battery (Korkman, Kirk, & Kemp, 1998), designed for neuropsychological testing of children. Standardised using 1200 healthy, representative US children (3-16 years) in 12 age bands, NEPSY II assesses a wide range of functions (attention and executive functioning, language, social perception, visuospatial processing, memory and learning, sensorimotor). However, not all subtests cover the full age range. Test-retest reliabilities were derived from a group of 165 children tested after 3 weeks and ranged from modest (.35) to very good (.94). Convergent validity with the D-KEFS in a sample of 49 children (9-16 years) was limited, as correlations were largely low to moderate (Korkman et al., 2007).

Verbal domain

Verbal fluency can be assessed in 5-16 year-olds using the NEPSY-II Word Generation subtest. It includes category and letter fluency, like the D-KEFS, and comparisons with design fluency (below) may be useful if there are domain specific problems. Test-retest reliabilities (13-16 years) were acceptable (.60 letter, .77 category).

NEPSY-II Animal Sorting is similar to the D-KEFS sorting task. This can, therefore, be regarded as a measure of planning, suitable for children 7-16 years. However, unlike the D-KEFS, perceptual and verbal sorts are not discriminated, the child is not asked to articulate their sorting strategy, and there are no explicitly verbal sorts (e.g. no word sorts). Test-retest reliabilities are reasonably good (.64 to .75).

Auditory Attention and Response Set includes a standardised assessment of inhibition in 7-16-year-olds, although only using the ‘response set’ part of the task. The test loosely resembles a Stroop task, hence is more verbal than visuospatial (stimulus items are verbal, but responses are via pointing). Test-retest reliabilities are moderate to good (.55 to .85, 7-12 years).

Inhibition (ages 5-16) requires naming of squares as circles and circles as squares (or up arrows as down arrows etc), combining verbal and visuospatial skills, but there is the facility to record behavioural observations such as uncorrected versus self-corrected errors, which may give an indication of subtle deficits not available in other batteries. Test-retest reliabilities vary from modest to good (5-16 years, range .35-.88). There is also a ‘switching’ part to this test (7-16-years), which has good reliability (.73-.90).

Visuospatial domain

Design Fluency is similar to the D-KEFS measure (see earlier), but is slightly simpler and covers a lower age range (5-12 years). Convergent validity data are not provided, and test-retest reliability was at best moderate (.44 to .63).

Clocks (7-16 years) tests planning and organisation, assessing a child's ability to tell the time both from traditional and digital clock faces, and to draw clock faces accurately. This task has less face validity than others, but reasonably good rest-retest reliabilities (.64-.82). Statue (3-6 years), which requires the child to remain motionless in a proscribed standing position despite distractions (e.g. a cough from the tester), is a measure of 'motor persistence' and inhibition. This is the only executive measure suitable for younger children, but is not a typical measure of inhibition. Nevertheless, test-retest reliabilities were good (.82 3/4-year-olds; .88 5/6-year-olds).

The advantage of NEPSY II is its breadth, useful for examining a range of abilities *including* executive functioning. The scoring software computes scaled scores and percentiles ranks by age and 'Behavioural Observations' supplement some subtests to provide helpful additional data. However, the full test is lengthy to administer (2/3 hours) and there are delayed subtests that require careful time management. More generally, many subtests do not yield pure measures of either verbal or visuospatial abilities (Design Fluency and Word Generation excepted), and there are fewer measures of executive functioning for children under 7 years.

Research using the original NEPSY to assess clinical groups (e.g. Riddle et al., 2005, with spina bifida; and Schmitt & Wodrich, 2004, with mixed neurological impairments) indicated that, whilst the battery had clinical validity, findings were somewhat mixed. Stinnet et al. (2002) and Ahmad and Warriner (2001) both provide critiques of the NEPSY. For example, Stinnet et al. (2002) suggest that the five original domains of assessment over-define the structure of the battery. Clearly, with the developments and improvements claimed for the NEPSY II, more research is required.

Other tests of executive functioning

Verbal working memory. The ‘listening span task’ measures executive-loaded working memory in the verbal domain (e.g. Siegel & Ryan, 1989; Leather & Henry, 1994). Participants listen to a series of sentences, decide whether each is true or false (e.g. ‘birds fly in the *sky* – true; ‘people live in *nests’* - false) and recall the final words of each sentence in order (sky, nests). A UK-normed, standardised version is available in the Working Memory Test Battery for Children (WMTB-C, Pickering & Gathercole, 2001) with an age range of 5-15 years. Although test-retest reliabilities are variable (.83 5-7 years, .38 9-11 years), Listening Recall is easy to administer and score, with clear procedures for continuing testing when threshold span is exceeded. [See also Automated Working Memory Assessment (AWMA, Alloway, 2007)].

Backward digit span, often regarded as a measure of executive-loaded working memory, is included in standard cognitive assessments, but rarely scored separately from forward digit span. The WMTB-C includes a separate measure of backward digit span with moderate to adequate test-retest reliability (.53 to .71), as well as a forward digit span measure. The WMTB-C also includes counting span, which requires visuospatial *and* verbal skills (counting dots on a series of cards and recalling the totals). It is easy to administer/score, with moderate to adequate test-retest reliabilities (.48. to 74).

Visuospatial working memory. In the ‘odd one out’ task (Henry, 2001; Hitch & MacAuley, 1991; Russell, Jarrold, & Henry, 1996), three similar (and not readily named) visual items are displayed in a left to right array. Children point to the item that is visually different from the other two and remember its spatial location. The task becomes more difficult when two or more ‘odd one out’ arrays are presented, before asking the participant to recall the spatial locations. This task is easy to administer/score, suitable from around 5 years and enjoyable, but no standardised version is available.

Inhibition. An engaging measure is the Animal-Stroop (Wright et al., 2003), developed for children 3-16 years. Stimuli include black-and-white cartoon style images (cow, pig, sheep, duck). In the congruent condition, animal images appear as normal and children name the *body* of the animal. In the incongruent condition, animal heads and bodies are jumbled up, but children *still* name the body of the animal. Stroop-like interference is caused by the preferential processing of facial information, which must be inhibited. This test has a wide age range, but combines visuospatial and verbal skills, making domain specific comparisons difficult. Reaction time and error rate measures have moderate to high reliabilities (.56 to .93) and although this is not a standardised test, the authors include comparison data on 155 healthy children in seven age bands.

Set shifting/switching. The Wisconsin Card Sorting Test (WCST) requires the matching of cards with very little instruction. Cards contain different representations of shape, colour and numerosity and are matched to one of four ‘base’ cards that vary along the same dimensions (e.g. one red triangle, two green stars, three yellow crosses, four blue circles). Feedback is given as to whether each match is correct, and matching dimensions change throughout the task. The measure of interest is success in switching from one matching dimension to another; poor executive skills result in perseveration. The WCST-64 (Kongs et al., 2000), suitable for individuals from 6-89 years, includes half the original cards. Test-retest reliability is reasonably good (.74) for adults (child data not available) and the WCST-64 correlates highly with the full test (.9). In validation studies, clinical groups of children (frontal lesions) performed worse than children with diffuse/non-frontal injuries as expected (Kongs et al., 2000). Although test presentation is straightforward, accurate *recording* is crucial and requires practice. The WCST is broadly in the visuospatial domain, although numerosity judgements may involve verbal counting.

Planning. Maze Tasks (Porteus Mazes, WISC-III, WPPSI-R, WMTB-C) can assess planning ability (Anderson, 1998); the child makes a line along a maze route without encountering ‘dead ends’ or going outside lines, requiring forward planning to avoid poor decision-making. Scoring difficulties among individuals with intellectual disabilities or developmental disorders can result from performance being confounded with fine motor skills, visual perception and speed of response (many maze tasks are timed). However, maze measures are readily available within working memory or IQ batteries, assess planning in the visuospatial domain and have UK-based norms and reasonably good reliabilities (e.g. WMTB-C Mazes task, test-retest reliability .68).

Interpreting tests of executive functioning

Tests of executive functioning must include novelty, complexity and the need to integrate information (Anderson, 1998; Shallice, 1990). Hence, low reliability estimates may be inevitable as, once a person has become familiar with a test, its ability to measure novelty is diminished. Therefore, it may be necessary to accept lower levels of reliability among these tests than is ideal. The choice of areas for assessment depends on the depth and detail required, as well as the difficulties experienced by the child. Tests are more limited for children under 8 years; two batteries do not cover this age range (D-KEFS, BADS-C) and coverage in a third is patchy (NEPSY II).

None of the test batteries described assess each of the five areas of executive functioning outlined at the beginning of the paper comprehensively, nor do most of the batteries provide comparative scores on verbal and visuospatial subdomains. The battery that comes closest to achieving comprehensive coverage and at least some comparisons between verbal and visuospatial domains is the D-KEFS. The CANTAB covers three executive areas, but only in the nonverbal domain and, given its hefty price, may not be the best value for money. The NEPSY-II quotes better reliabilities than some of the other batteries, yet whilst it is reasonably comprehensive on first sight, many individual subtests have restrictive age ranges. The BADS-C offers the clinician ecological validity for assessing planning (plus one test of set shifting), yet many of the reliabilities are perhaps worryingly low, even taking into account the point made above regarding reliabilities. Perhaps the most consistently reliable test is the TEA-Ch, but it is provides mainly measures of inhibition. Overall, in choosing appropriate test batteries, clinicians will need to pinpoint the *types* and *range* of executive difficulties they want to assess (this might depend on the client group), the relevant age range, whether comparisons between visuospatial and verbal processing are required, and whether there are verbal or motor difficulties. At present, to assess executive functioning thoroughly, a combination of tests will be necessary, possibly including both a test battery and some of the ‘other’ measures of executive functioning outlined earlier.

Measures of executive functioning focus on cognitive processes in one-to-one situations, and clinicians may also require alternative assessments (observation, interviews with families/schools) to understand how executive dysfunction affects emotional responses (e.g. motivation), cognitive abilities and behavioural actions (e.g. socially inappropriate behaviour) in everyday life (Anderson, 2002). NEPSY II provides for recording behavioural observations in some tests. Additional behavioural evidence may be obtained using the BRIEF (Behavior Rating Inventory of Executive Function, Gioia et al., 2000), which assesses EF in the home and school environments (5- to 18-year-olds) using parent and teacher questionnaires (test-retest reliability is good, 80-.98.). (See also DEX-C from BADS-C).

Summary

No single battery of executive functioning offers a complete assessment of executive functioning in school-age children in both the visuospatial and verbal domains; a combination of batteries and stand-alone tests may offer maximum clinical flexibility. Choice of tests will depend on how important ease of comparison across different domains is for particular clients, as well as whether there are associated difficulties, for example, with motor skills or verbal expression.

Acknowledgements

The preparation of this paper was supported by an award from the ESRC UK to the first author (Grant reference: RES-062-23-0535).

Table 1. Summary of test EF subcomponents and reliabilities data

Measure	Age range	Test/retest *r*	Sample size	EF domain
BADS-C	7-16 years		259	
Playing Cards		-0.24		Mental Flexibility
Water Test		*Not available*		Planning
Key Search Test		0.81		Planning
Zoo Map		0.29-0.59		Planning
Six Part Test		0.44		Planning
CANTAB	4-16 years		2000††	
Stop Signal Task		*Not available*		Inhibition
IED		0.40		Mental Flexibility
Stockings of Cambridge		0.64		Planning
D-KEFS	8-16 years		1750	
Switching		0.13		Mental Flexibility
Stroop		0.77-0.90		Inhibition
Trail Making		0.20-0.82		Mental Flexibility
Tower Task		0.51		Planning
Sorting Test		0.55-0.80		Planning
Design Fluency		0.43-0.66		Fluency
Verbal Fluency		0.37-0.80		Fluency
NEPSY-II	3-16 years†		1200	
Word Generation		.60-.77		Fluency
Animal Sorting		.64-.75		Planning
Auditory Attention & Response Set		.55-.85		Inhibition
Inhibition		.73-.90		Inhibition
Design Fluency		.44-.63		Fluency
Clocks		.82-.88		Inhibition
TEA-Ch	6-16 years		293	
Walk Don't Walk		0.73		Inhibition
Opposite Worlds		0.92		Inhibition

Creature Counting		0.69-0.73		Mental Flexibility
Sky Search		0.81		Dual Task Interference
WMTB-C	4-15 years		750	
Listening Recall		0.38-0.83		Working Memory
Backward Digit Span		0.53-0.71		Working Memory
Counting Span		0.48-0.74		Working Memory
Animal Stroop	3-16 years	0.56-0.93	155	Inhibition
WCST-64	6.5-17 years	0.74	452	Mental Flexibility

† not all subtests suitable for entire age range – see text for details

†† sample is mixed adult and children

References

Ahmad, S. A., & Warriner, E.M. (2001). Review of the NEPSY: A developmental neuropsychological assessment. *Clinical Neuropsychologist, 15*, 240-249.

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

Anderson, V. (1998). Assessing executive functions in children: Biological, psychological and developmental considerations. *Neuropsychological Rehabilitation, 8*, 319-349.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*, 71-82.

Baddeley, A.D. (1986). *Working memory*. Oxford: OUP.

Baddeley, A.D. (1997). *Human memory: Theory and practice*, (Revised Edition) East Sussex: Psychology Press Ltd.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive sciences, 4*, 417-423.

Baddeley, A.D. (2007). *Working memory, thought, and action*. Oxford: OUP.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), *The psychology of learning and motivation, Vol. 8* (pp. 47-89). New York: Academic Press.

Barkley, R. A. (1997). Behavioural inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*, 65-94.

Bishop, D.V.M., & Norbury, C.F. (2005a). Executive functions in children with communication impairments, in relation to autistic symptomatology 1: Generativity. *Autism, 9*, 7-27.

Bishop, D.V.M., & Norbury, C.F. (2005b). Executive functions in children with communication impairments, in relation to autistic symptomatology 2: Response inhibition. *Autism*, *9*, 29-43.

Cambridge Cognition Ltd (2006a). *Cambridge Neuropsychological Test Automated Battery* (CANTAB). Cambridge: Cambridge Cognition Ltd.

Cambridge Cognition Ltd (2006b). *Cambridge Neuropsychological Test Automated Battery (CANTAB): Test-Retest Reliability Characteristics.* Cambridge: Cambridge Cognition Ltd.

Channon, S., Pratt, P., & Robertson, M.M. (2003). Executive function, memory, and learning in Tourette's syndrome. *Neuropsychology*, *17,* 247-254.

Curtis, W.J., Lindeke, L.L., & Georgieff M.K. (2002). Neurobehavioural functioning in neonatal intensive care unit graduates in late childhood and early adolescence, *Brain, 125,* 1646–1659.

Delis, D.C., Freeland, J., Kramer, J.H., & Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: Construct validation of the California Verbal Learning Test. *Journal of Consulting and Clinical Psychology*, *56,* 123-130.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Function System, Technical Manual* (D-KEFS). San Antonio: The Psychological Corporation.

Delis, D.C., Kramer, J.H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: An update. *Journal of the International Neuropsychological Society, 10*, 301-303.

Diamond, A., & Goldman-Rakic, P. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral pre-frontal cortex. *Experimental Brain Research, 74*, 24-40.

Ellis Weismer, S., Evans, J., & Hesketh, L.J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech Language and Hearing Research, 42*, 1249-1260.

Emslie, H., Wilson, F., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003). *Behavioural Assessment of the Dysexecutive Syndrome for Children (*BADS-C*).* London: Harcourt Assessment/The Psychological Corporation

Fisk, J.E., & Sharp, C.A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology, 26,* 874-890.

Garon, N., Bryson, S.E., & Smith, I.M. (2008). Executive function in pre-schoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.

Guerts, H.M., Verte, 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,* 836-854.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function (BRIEF).* Lutz, FL: Psychological Assessment Resources.

Henry, L.A. (2001). How does the severity of a learning disability affect working memory performance? *Memory, 9*, 233-247.

Hill, E.L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24,* 189-233.

Hitch, G.J., & McAuley, E. (1991). Working memory in children with specific arithmetic learning difficulties. *British Journal of Psychology, 82*, 375-386.

Homack, S., Lee, D., & Riccio, C. (2005). Test review: Delis-Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology, 27*, 599-609.

Hughes, C., Plumet, M.H., & Leboyer, M., (1999). Towards a cognitive phenotype for autism: Increased prevalence of executive dysfunction and superior spatial span amongst siblings of children with autism. *Journal of Child Psychology and Psychiatry, 40*, 705-718.

Huizinga, M., Dolan, C., & van der Molen, M. (2006). Age related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017-2036.

Jones-Gotman, M., & Milner, B. (1977). Design fluency: The invention of nonsense drawings after focal cortical lesions. *Neuropsychologia, 15*, 653-674.

Kongs, S.K., Thompson, L.L., Iverson, G.L., & Heaton, R.K. (2000). *Wisconsin Card Sorting Test - 64 Card Version* (WCST-64). Odessa, Florida: Psychological Assessment Resources, Inc.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY*: *A developmental neuropsychological assessment manual.* San Antonio: Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY II*: *A developmental neuropsychological assessment manual.* San Antonio: Psychological Corporation.

Lanfranchi, S., Cornoldi, C., & Vianello, R. (2004). Verbal and visuospatial working memory deficits in children with Down syndrome. *American Journal of Mental Retardation, 109*, 456-466.

Leather, C., & Henry, L.A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58,* 88-111.

Lehto, J.E. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology, 49A*, 29-50.

Lehto, J.E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59-80.

Levin, H.S., Culhane, K.A., Hartmann, J., Evankovich, K., Mattson, A.J., Harward, H., Ringholz, G., Ewing-Cobbs L., & Fletcher, J.M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 7*, 377-395.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, 42,* 261-270.

Luciana, M. (2003). Practitioner Review: Computerised assessment of neuropsychological function in children: Clinical and research applications of the Cambridge Neuropsychological Testing Automated Battery (CANTAB). *Journal of Child Psychology and Psychiatry, 44*, 649-663.

Luciana, M., & Nelson, C.A., (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia, 36***,** 273-293.

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The Test of Everyday Attention for Children* (TEA-Ch). Bury St Edmunds, UK: Thames Valley Test Company.

Manly, T., Anderson, V, Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I.H. (2001). The differential assessment of children's attention: The Test of Everyday Attention for Children (TEA -Ch), normative sample and ADHD performance. *Journal of Child Psychology and Psychiatry, 42*, 1065-1081.

Martin, A., Wiggs, C.L., Lalonde, F. & Mack, C. (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks, *Neuropsychologia, 32*, 1487-1494.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J.M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 313-334). New York: McGraw-Hill.

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’ to tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R., Klin, A., McMahon, W., Minshew, N., Munson, J., Pennington, B., Rogers, S., Spence, M., Tager-Flusberg, H., Volkmar, F., & Wrathall, D. (2004). Performance on Cambridge Neuropsychological Text Automated Battery subtests sensitive to frontal lobe function in people with autistic disorder: Evidence from the Collaborative Programs of Excellence in Autism network. *Journal of Autism and Developmental Disorders, 34*, 139-150.

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

Pennington, B.F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37,* 51-87.

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children (WMTB-C).* London: The Psychological Corporation.

Reitan, R.M., & Wolfson, D. (1992). *Neuropsychological evaluation of older children.* Tuscan, AZ: Neuropsychology Press.

Riddle, R., Morton, A., Sampson, J.D., Vachha, B., & Adams, R. (2005). Performance on the NEPSY among children with spina bifida. *Archives of Clinical Neuropsychology, 20*, 243-248.

Robinson, S., Goddard, L., Dritschel, B., Wisley, M. & Howlin, P. (2009). Executive functions in children with autism spectrum disorders. *Brain and Cognition, 71*, 362-368.

Russell, J., Jarrold, C., & Henry, L. (1996). Working memory in children with autism and with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 37,* 673-686.

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*, 745-759.

Schmitt, A.J., & Wodrich, D.L. (2004). Validation of a Developmental Neuropsychological Assessment (NEPSY) through comparison of neurological, scholastic concerns, and control groups. *Archives of Clinical Neuropsychology, 19*, 1077-1093.

Schwartz, S., & Baldo, J., Graves, R., & Brugger, P. (2003). Pervasive influence of semantics in letter and category fluency: A multidimensional approach. *Brain and Language, 87,* 400-411.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled. *Child Development, 60,* 973-980.

Shallice, T. (1990). *From neuropsychology to mental structure*. New York: Cambridge University Press.

Stinnet, T.A., Oehler-Stinnet, J., Fuquq, D.R., & Palmer, L.S. (2002). Examination of the underlying structure of the NEPSY: A developmental neuropsychological Assessment. *Journal of Psychoeducational Assessment, 20,* 66-82.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 634-662.

Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Turner, M.A. (1999). Generating novel ideas: Fluency performance in high-functioning and learning disabled individuals with autism. *Journal of Child Psychology and Psychiatry, 40*, 189-201.

Welsh, M.C., Pennington, B.F., & Groisser, D.B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7,* 131-149.

Wilding, J. (2005). Is attention impaired in ADHD? *British Journal of Developmental Psychology, 23*, 487-505.

Williams, B.R., Ponesse, J.S., Schachar, R. J., Logan, G.D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology, 35*, 205-213.

Wright, I., Waterman, M., Prestcott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: Typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*, 561-557.