Journal of Experimental Child Psychology 138 (2015) 31–53

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Cognitive flexibility in young children: General or task-specific capacity?

Gedeon O. Deák [a,*], Melody Wiseheart [b]

[a] *Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093, USA*
[b] *Department of Psychology and LaMarsh Centre for Child and Youth Research, York University, Toronto, Ontario M3J 1P3, Canada*

ARTICLE INFO

Article history:
Received 25 June 2014
Revised 8 April 2015

Keywords:
Causal reasoning
Cognitive flexibility
Executive functions
Individual differences
Inhibition
Rule switching
Word learning
Working memory

ABSTRACT

Cognitive flexibility is the ability to adapt to changing tasks or problems. To test whether cognitive flexibility is a coherent cognitive capacity in young children, we tested 3- to 5-year-olds' performance on two forms of task switching, rule-based (Three Dimension Changes Card Sorting, 3DCCS) and inductive (Flexible Induction of Meaning–Animates and Objects, FIM-Ob and FIM-An), as well as tests of response speed, verbal working memory, inhibition, and reasoning. Results suggest that cognitive flexibility is not a globally coherent trait; only the two inductive word-meaning (FIM) tests showed high inter-test coherence. Task- and knowledge-specific factors also determine children's flexibility in a given test. Response speed, vocabulary size, and causal reasoning skills further predicted individual and age differences in flexibility, although they did not have the same predictive relation with all three flexibility tests.

Introduction

Cognitive flexibility is the capacity to modify working memory, attention, and response selection in response to changing endogenous and exogenous task demands. Cognitive flexibility has been the focus of behavioral and neuropsychological studies (e.g., Eslinger & Grattan, 1993; Kramer, Cepeda, & Cepeda, 2001; Smith & Blankenship, 1991) using a variety of tasks and contexts and wide age ranges

* Corresponding author.
E-mail address: gdeak@ucsd.edu (G.O. Deák).

http://dx.doi.org/10.1016/j.jecp.2015.04.003
0022-0965/

(Ionescu, 2012). Age-related changes in cognitive flexibility have been reported in tests of rule switching (Zelazo, Frye, & Rapus, 1996), word learning (Deák, 2003), spatial reasoning (Hermer-Vazquez, Moffet, & Munkholm, 2001), categorization (Blaye & Bonthoux, 2001), and problem solving (Chen, 1999). Many studies and paradigms suggest that flexibility improves significantly from 3 to 6 years of age. If flexibility develops similarly across multiple tasks, it might mean that flexibility is a generalized cognitive capacity—an "executive" control process that operates over a wide range of task contexts (e.g., Martin & Rubin, 1995; Zelazo & Frye, 1998).

The idea of general cognitive capacities has a long history in psychology (e.g., Ackerman, 1988; Engle & Kane, 2004; Humphreys, 1979). Many researchers have argued that a few general *executive functions* (EFs) control cognition in a variety of tasks and contexts (but see Barkley, 2012; Jurado & Rosselli, 2007). Many proposed EF frameworks incorporate a function of cognitive flexibility or "set shifting" (e.g., Miyake et al., 2000). A related hypothesis is that EFs are stable endogenous traits of individuals (Friedman et al., 2008). This implies that individual differences in cognitive flexibility should be constant across tasks, times, and content. Some authors have suggested that these general EFs, including flexibility, mature and stabilize during early childhood (Carlson, Moses, & Breton, 2002; Davidson, Amso, Anderson, & Diamond, 2006).

That hypothesis is controversial; an alternative is that flexibility develops in a domain-specific fashion as children gain task-specific skills and knowledge (Luwel, Verschaffel, Onghena, & De Corte, 2003; Ravizza & Carter, 2008). By this view, flexibility might improve in many tasks between 3 and 5 years of age simply because children acquire a great deal of varied knowledge and skills during that time. That is, flexibility might improve due to parallel gains in knowledge and skills across domains, not to the development of a generalized EF. If this is true, older children's flexibility should relate to individual domain-specific skills. For example, it has been shown that school-aged children's flexibility in reading-related tasks is partly predicted by their reading skill (Cartwright, Marshall, Dandy, & Isaac, 2010).

It is also possible that children's flexibility is determined by both a general EF and task- or domain-specific skills and knowledge. Another related possibility is that there are several dissociable, moderately general flexibility capacities, and each is more relevant to (or more heavily recruited for) some tasks than others (Kim, Johnson, Cilles, & Gold, 2011). Both of these alternatives would predict limited between-test intra-individual coherence of flexibility.

Determining whether children's cognitive flexibility depends on general capacities, on task-specific knowledge and skills, or on both would go some way towards explaining developmental changes in cognitive control. However, there is little evidence concerning the coherence of children's flexibility. Most studies implicitly treat flexibility as a general capacity that can be assessed by a single rule-switching test despite the fact that external validity and construct validity of most tests has not been established.

To address this question, we gave preschool children three tests of flexibility representing two types of cognitive skills or domains. If individual children's flexibility is similar across all tests, it will imply a general capacity. If it is consistent only between two tests from the same task domain, it will suggest that flexibility is determined by task-specific skills, or by several moderately specific capacities, or both. If flexibility is inconsistent across all three tests, it will suggest that flexibility is largely determined by task-specific knowledge.

Selecting comparable tests with different content domains and task demands is challenging because most studies of young children use one test, the Dimensional Change Card Sorting test or DCCS (Zelazo, 2006). This is a *rule-switching* test; children learn two deductive binary rules for sorting two stimuli. They are told to follow one rule and, at some later time, to switch to the other rule. The test yields robust age differences; most 3-year-olds fail to follow an instruction to switch to the second rule, but most 5-year-olds correctly switch. The test classifies each child as flexible or inflexible with little further differentiation. Although recent studies have explored more sensitive measures of rule-switching efficiency in older children (e.g., Cepeda, Kramer, & Gonzalez de Sather, 2001), these paradigms are not well-suited for preschool children.

Other researchers have, however, tested preschoolers using age-appropriate tests that yield parametric estimates of flexibility. These tests involve more subtasks and switches, as well as more trials and response options, than the DCCS (Deák & Narasimham, 2003, 2014; Narasimham, Deák, &

Wiseheart, 2015). Notably, the tests also represent a different type of task, *cue induction*, rather than rule switching. Cue induction is the common process of selecting and integrating multiple sources of information that are probabilistically related to some task or judgment; such a judgment is inductive (i.e., indeterminate). Cue-induction flexibility is needed for making different inferences based on different subsets of available information or cues.

Cue-induction tests of flexibility are useful because rule-switching tests might not capture young children's common everyday cognitive activities. Rule-switching tests demand arbitrary reversals of symbolic mappings, which play a small role in preschoolers' everyday experience (Deák, 2003; see also Burgess et al., 2006). These reversals are analogous to solving an algebra problem with the premises "Let $x = 4$ and $y = 3$" then getting another problem with the (switched) premises "Let $x = 3$ and $y = 4$." Such arbitrary mapping reversals are an unusual sort of symbol manipulation; in fact, they are confusing for adolescents learning algebra (Knuth, Alibali, McNeil, Weinberg, & Stephens, 2011). If these reversals are unfamiliar to preschoolers, rule-switching tests might be assessing a fairly peculiar skill, not one that generalizes to everyday tasks that require flexibility. This might explain why brief feedback or practice can eliminate preschoolers' switching errors (Bohlmann & Fenson, 2005; Perner & Lang, 2002). It can also explain why rule switching improves from 3 to 5 years, an interval when many children start attending preschool classes that impose an expanding, increasingly elaborate schedule of rules.

If cognitive flexibility reflects task-specific skills rather than a generalized EF, rule switching might be an acquired skill—a learned ability to process, adopt, and reverse arbitrary rule-to-response mappings. However, many everyday situations instead require children to shift attention and modify behavior in response to probabilistic social or linguistic cues that are associated with the prevailing task context. These social and linguistic cues seldom reverse or change arbitrarily; instead, new cues are usually related to some social event (e.g., topic shift, new interlocutor, new information). In addition, the cues are seldom explicitly stated or explained. Thus, everyday flexible cue induction requires sensitivity to changeable, probabilistic, implicit, and pragmatically constrained contextual information.

Preschool children can flexibly use such cues to make inductive judgments (e.g., Nguyen & Murphy, 2003). For example, between 3 and 5 years of age, children become more flexible at using changing semantic cues to infer novel word meanings (Deák, 2000, 2003; Deák & Narasimham, 2003, 2014). When told that an object is "made of molap," most preschoolers infer that *molap* refers to its material. Later, when told that the same object "has a fodi," most children will infer that *fodi* refers to a salient part, not its material. These inferences require children to constrain the possible meanings of successive words according to each one's specific semantic context. This paradigm encapsulates a pervasive demand of children's language learning: flexibly using implicit cues to interpret unfamiliar words.

Cue-induction flexibility improves from 3 to 5 years of age, parallel to improvement in rule-switching flexibility. This parallel development might suggest a generalized capacity for flexibility. Alternately, it might be circumstantial, given that most cognitive tests show improvement from 3 to 5 years. Suggestively, there is evidence that cue-induction flexibility and rule-switching flexibility rely on distinct neural substrates. Studies of adult humans and rats suggest a partial dissociation between hippocampal mechanisms for learning specific, well-defined contingencies (i.e., rules) and striatal mechanisms for learning probabilistic cue–outcome associations (Frank, O'Reilly, & Curran, 2006; O'Reilly & Frank, 2006). The former might contribute more to rule-switching flexibility and the latter to cue-induction flexibility (Thompson-Schill, Ramscar, & Evangelia, 2009). Both develop during early childhood (Ramscar, Dye, Gustafson, & Klein, 2013), but no study has directly compared children's rule- switching flexibility and cue-induction flexibility.

One issue to consider in comparing flexibility across tests is *subtask difficulty*—that is, the difficulty (based on discriminability, specificity, etc.) of specific cues for each problem type within a test. Children's ability to comprehend and use a particular cue or rule will affect their performance on specific questions or subtasks, and their overall flexibility on the test. Errors (e.g., perseverating on a rule) might be due not to general inflexibility but rather to poor comprehension of a cue or rule. For example, 3-year-olds' comprehension of words in the DCCS (e.g., "color," "shape") predicts whether they perseverate (Munakata & Yerys, 2001). In addition, the strength of preschoolers' working memory representation of rules determines their rule-switching speed even when they do not make errors (Holt & Deák, 2015). Children's conceptual knowledge also affects how readily they switch between subtasks (Blaye, Bernard-Peyron, Paour, & Bonthoux, 2006; Deák, Ray, & Pick, 2004). Thus, specific knowledge affects flexibility, and it is important to control or assess the specific difficulty of each (cue- or rule-based) subtask, and of each test overall, in order to interpret similarities and differences between tests.

To minimize this problem, we applied several strategies in our design. First, tests were designed to be similar in difficulty so that differences in flexibility would not be entirely due to between-test differences in subtask difficulty. However, it can be challenging to equate cue or rule difficulty across tests for young children. Thus, a second strategy was to make subtasks within each test sequenced similarly, starting with the easiest cue/rule first, then the next harder cue/rule, and finally the hardest cue/rule. This is necessary because subtask order can affect flexibility (e.g., Deák, Ray, & Pick, 2004; Ellefson, Shapiro, & Chater, 2006). If order effects were inconsistent across tests, it would complicate or invalidate between-test comparisons. Third, we assessed flexibility in each test using a measure that partly corrects for differences in subtask difficulty, [Correct Switches] ÷ [Opportunities to Switch] or CORSWOPS (see below). Fourth, in some analyses each child's accuracy on the first cue/rule of each flexibility test was treated as a covariate. This separates some variance due to task-related cue/rule comprehension. Finally, children completed tests of conceptual and linguistic knowledge (e.g., receptive language; see below) to determine whether these factors predicted test-by-test variability in flexibility.

The last strategy also supports a secondary goal of this investigation—to determine whether flexibility (rule switching, cue induction, or both) relates to children's EFs. Many EFs change greatly from 3 to 5 years of age while cognitive flexibility is developing. Perhaps changes in EFs contribute to changes in flexibility (Davidson et al., 2006; Thompson-Schill et al., 2009). Miyake and colleagues (2000) argued that flexibility is distinct from, but related to, other EFs, including working memory and cognitive inhibition. It has been suggested that children's flexibility might also be related to EFs—to cognitive inhibition (Zelazo, Müller, Frye, & Marcovitch, 2003), to inhibition and working memory (e.g., Carlson, 2005), and/or to processing speed (Cepeda, Cepeda, & Kramer, 2000). Currently, the relation remains unclear. Most previous studies have tested only rule-switching flexibility, so the relation of EFs to cue-induction flexibility as well as rule-switching flexibility has not been explored. However, two studies have found no reliable relation between cue-induction flexibility and verbal inhibition (Deák & Narasimham, 2003, 2014). The current study examined relations between both types of flexibility and three EFs: working memory, inhibition, and processing speed.

In sum, this study addressed three main questions. First, we investigated whether there is a generalized capacity for flexibility, as indicated by within-child consistency between rule-switching and cue-induction tests. A finding of consistency between two tests of the same type (e.g., cue-induction flexibility), but not with one of another type (e.g., rule-following flexibility), would suggest distinct task-related flexibility mechanisms but no global capacity. Second, we investigated whether shared variance in flexibility could be attributable to linguistic and conceptual knowledge. If flexibility is predicted by receptive vocabulary, for example, it would imply that cue comprehension mediates cognitive flexibility. Third, we investigated whether three EFs—working memory, inhibition, and processing speed—predict children's flexibility across tests.

To assess flexibility, 3- and 4-year-olds completed a test of rule-following flexibility and two tests of cue-induction (word-meaning) flexibility.[1] All tests provided parametric and nonparametric measures of flexibility because, unlike binary two-alternative forced-choice tests (e.g., DCCS), each test switched among three rules or cues with larger sets of more complex stimuli and more response options. These features provide more test sensitivity and more differentiated responses (e.g., both perseverative and haphazard errors; see Barceló & Knight, 2002).

The parametric rule-switching test was the Three Dimension Changes Card Sorting (3DCCS) test, which uses three sorting rules—size, color, and shape—and two rule switches (Cepeda & Munakata, 2007; Deák, 2003). This requires more complex stimuli than the DCCS. Children sort test cards with four different values for each of three properties, as in Fig. 1. Thus, there are four possible sorting

[1] Ideally, children would have completed two tests of each type; however, only one test of rule-following flexibility was available that yields parametric estimates of flexibility in children as young as 3 years.

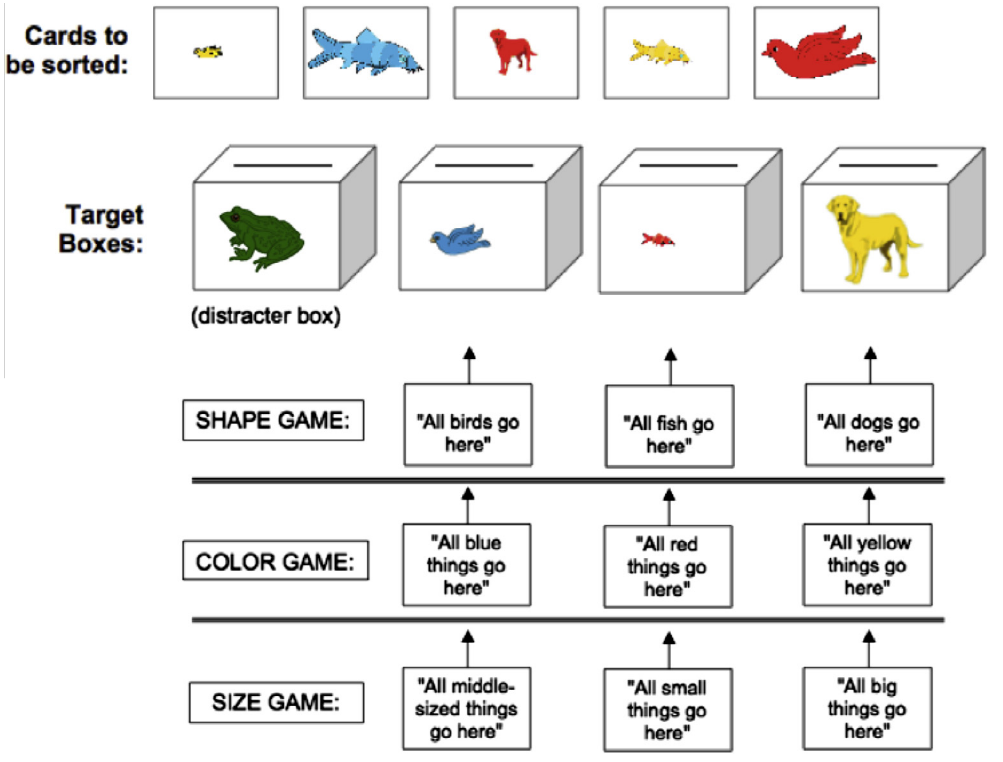

Fig. 1. Sample 3DCCS stimuli. Test cards (to be sorted) varied in shape (i.e., animal), color, and size. Each card could be sorted in a different box with a distinct target card, depending on the current rule (i.e., game): "shape game," "color game," or "size game." (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

responses (boxes) per trial. The test can yield both perseveration and haphazard-switching errors. However, overall flexibility in the 3DCCS is strongly correlated with flexibility in the DCCS (Narasimham et al., 2015), indicating convergent validity between the tests.

Children's cue-induction flexibility was measured in two FIM (Flexible Induction of Meaning) tests. The FIM–Objects (FIM-Ob) test (Deák, 2000) presents words for object properties. Children hear three novel words for the same novel objects (see Fig. 2). Each word can refer to one of three properties: shape, material, or a part. The three words for each standard object follow three different phrase cues: "is made of," "is shaped like a(n)," and "has a(n)." Children must infer each word's referent property and identify another object with that property. Because the cue and word change on each trial with a given set, children should generalize each word to a different property.

The FIM-Ob test reveals robust age and individual differences in flexible use of cues for word learning (Deák, 2000, 2003). Few 3-year-olds flexibly use phrase cues to infer different meanings, whereas most 4-year-olds and nearly all 5-year-olds do so. Variability across this age range is comparable to the 3DCCS. This allows us to assess between-test similarities in individual flexibility. Any similarities can be "triangulated" by comparing each test with a third test.

The other FIM test presents words for properties of animate creatures or FIM-An (Deák & Narasimham, 2014). Children hear three novel words for pictorial stimuli, each showing a creature in an alien environment holding a novel object (Fig. 3). The novel words for each standard follow, on different trials, three different phrase cues: "is a(n)," "lives in/on a(n)," and "holds a(n)." Again, children can use the cues to infer each word's referent and identify another picture with that property. This tests a similar kind of flexibility as the FIM-Ob but with different stimulus categories, materials,

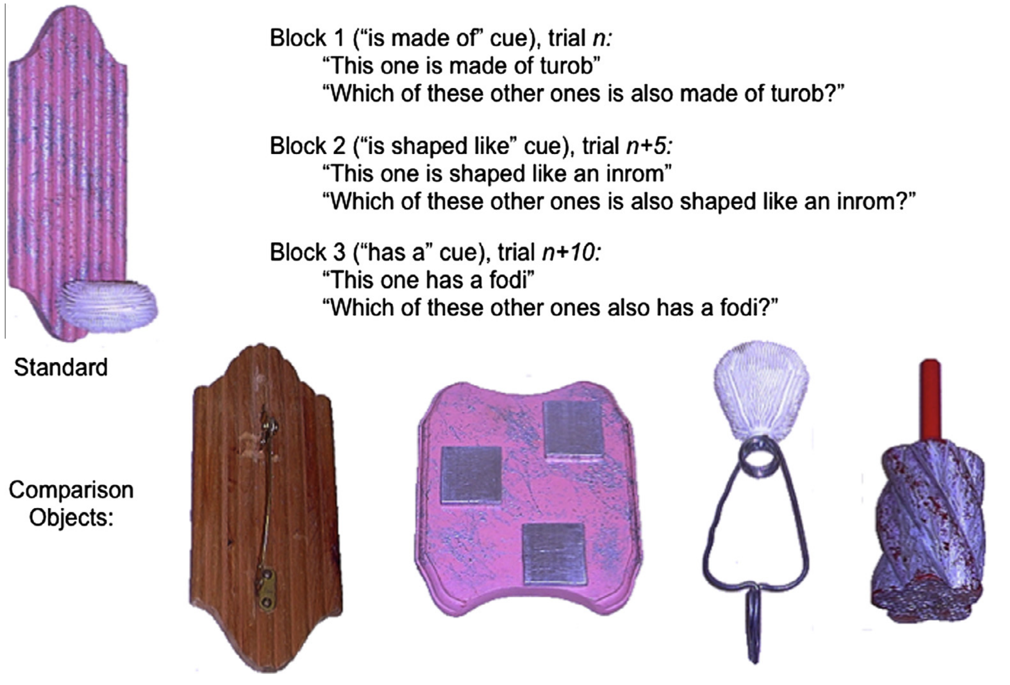

Fig. 2. One of five FIM-Ob test sets with example prompts from 3 trials. Top left object: standard. Comparison objects (bottom) from left: same shape, same material, same part, and distracter. Blocks include 5 trials, 1 per set, with the same cue. In the example, *turob* would generalize to the object second from left, *inrom* to the left-most object, and *fodi* to the object second from right.

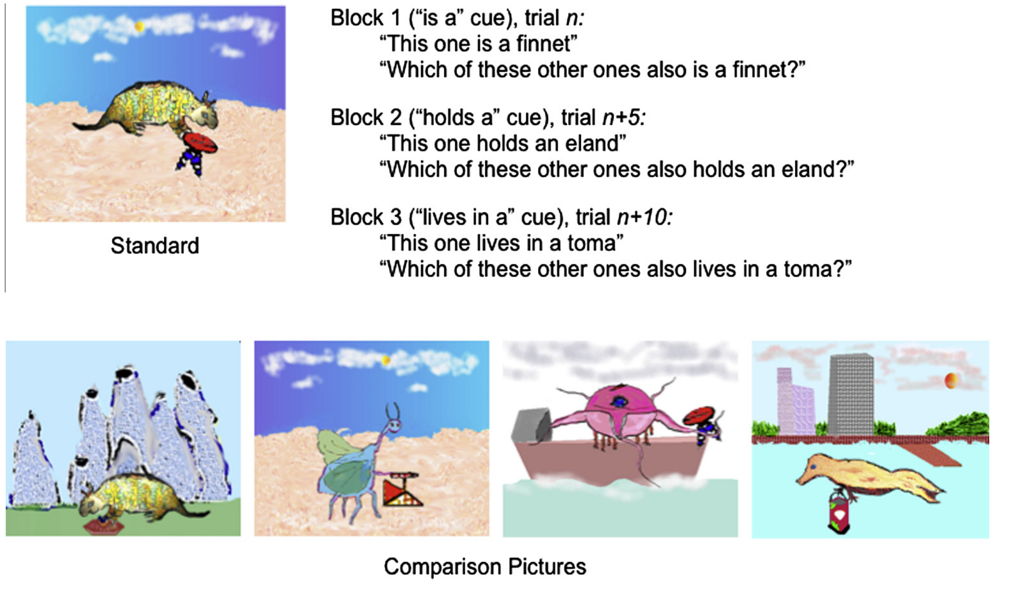

Fig. 3. One of five FIM-An test sets with example prompts from 3 trials. Top left image: standard. Comparison items (bottom) from left: same species, same habitat, same possession, and distracter. Blocks include 5 trials, 1 per set, with the same cue. In the example, *finnet, toma,* and *eland* would generalize to the first, second, and third items from left, respectively.

cues, and properties. Preschoolers' flexibility in the FIM-Ob and their flexibility in the FIM-An are moderately strongly correlated (Deák & Narasimham, 2014) even with age and receptive vocabulary controlled. The current study attempted to replicate that finding with minor procedural modifications.

Although the FIM-An and FIM-Ob both are cue-induction tests and the 3DCCS is a rule-switching test, the FIM-An and 3DCCS share other features: their stimuli come from the same domain (biological kinds) and share the same medium (colored pictures). These similarities might contribute to between-test associations. However, all tests differ in specific cues/rules, stimuli, and properties, so if they are correlated it could imply a general cognitive trait.

All three tests require receptive language ability to process cues or rules. This ability varies across children, so participants completed the Peabody Picture Vocabulary Test (PPVT), a normed receptive language test (Dunn & Dunn, 1997). To assess whether conceptual knowledge predicts flexibility, we selected an age-appropriate test of conceptual knowledge. Das Gupta and Bryant (1989) showed children object transformations of varying typicality and asked them to select the likely instrument of transformation. Although it directly assesses only a narrow range of conceptual content, this test might assess variance in conceptual knowledge more broadly. In addition, children's accuracy in the first block of each flexibility test provides converging evidence of their task-relevant linguistic and conceptual knowledge.

Children completed several EF tests. Based on a hypothesis that task switching demands inhibitory processes (Miyake et al., 2000; Zelazo et al., 2003), children completed two age-appropriate tests of cognitive and behavioral inhibition: one that requires inhibiting strong verbal associations, the Stroop Day–Night test (Gerstadt, Hong, & Diamond, 1994), and one that requires inhibiting an imitative tendency, Luria's Tapping test (Luria, 1962/1966). In addition, based on models that task switching requires working memory activation and maintenance of the current rule (e.g., Baddeley, Chincotta, & Adlam, 2001; Cepeda et al., 2000), children completed a test of verbal working memory (vWM), Memory for Names (Woodcock & Johnson, 1989). Finally, processing speed is a task-general individual difference that modulates cognitive control (Kail, 1991; Kail & Hall, 1994); here we considered it an EF parameter. Response speed was assessed with the Box Completion test (Salthouse, 1994), which can be administered to young children.

Although these measures barely tap the range of cognitive capacities that might relate to cognitive flexibility,[2] they serve as a starting point; if any are consistently associated with the flexibility tests, they will suggest relations that merit further investigation.

Method

Participants

A total of 93 3- and 4-year-olds were recruited from local preschools, and 85 completed all three sessions (8 children were excluded due to absence or refusal to participate in one or more sessions). In addition, a replication group of 12 3- and 4-year-olds was recruited after the main study to test for the possibility of order effects; these children completed all tests in a different order. Two children did not complete all sessions, leaving 10 children (6 girls) in the replication group (5 3-year-olds and 5 4-year-olds). Extensive comparisons for differences between the main group and the replication group revealed almost identical performance on all tests. Thus, their data were pooled and analyzed as a single group of $N = 95$ children[3] (47 girls, mean age = 49 months, range = 36–59). Children were tested in their preschool. All procedures were approved by the university's institutional review board.

[2] We assessed one other capacity, namely children's awareness of indeterminacy (Klahr & Chen, 2003). Deák and Enright (2006) found that this was correlated with children's ability to switch answers to similar but distinct questions. Therefore, we administered an expanded set of questions like those in Deák and Enright (2006). However, the results showed a large floor effect, suggesting that the expanded test was too difficult. Thus, the results, unfortunately, are not and will not be considered further.

[3] This does not affect any of the results below; it simply increases statistical power. Nonetheless, details of group performance are available from the corresponding author. The reason for this design was that if there were order effects, randomizing test order would have rendered the data ambiguous. A consistent test order allowed comparison of individual differences. The replication group provided a check for order effects that could have limited the interpretability of the results.

Most parents (79%) completed and returned a questionnaire about family demographics and child history. (Children whose parents did not return it performed no differently on any test than children whose parents did return it.) Children's ethnic distribution was 7% Asian, 23% multiracial or "other", 3% Hispanic, and 68% White and non-Hispanic. Parents' mean age was 39 years (*SD* = 4), with 17 years of education (*SD* = 2). Most children (89%) lived with two caregivers. Most children (61%) had one sibling, 13% were singletons, and 25% had two or more siblings. No child had any known sensory or cognitive problems except one child with corrected vision. Mean gestational age was –0.6 weeks from term (*SD* = 1.8), and birth weight (reported for only 58 children) averaged 3.3 kg (*SD* = 0.6), comparable to the U.S. median in 2012 (3.25 kg; Centers for Disease Control., 2013). One child had a (minor) birth complication. All children spoke English fluently, and 38% had exposure to a second language. PPVT-3 scores indicated that children with second-language exposure did not differ from those without it (second-language exposure mean = 108.4, *SD* = 11.7; English-only mean = 109.3, *SD* = 11.0).

Overall design and procedure

Children participated in a quiet room in their preschool. Each child completed three sessions (typically 30–45 min long) within a 2-week period. Tests were presented in the same order to all children to avoid order effects. The orders for the main and replication samples are shown in Appendix A. Orders were quasi-randomly determined with the following constraints. Flexibility tests were administered in different sessions. The replication order was also constrained such that each test was switched to a new session, in a different ordinal position, with different preceding and following tests. (Another test not reported here measured children's tool-using flexibility. There were no strong associations between it and the other flexibility tests. For that reason, and because it was rather elaborate, it will be described in Deák & Boddupalli, 2015).

One concern was that children might respond similarly across flexibility tests if the testing situation primes response strategies from the previous test session(s). Any such between-session situational priming could spuriously increase between-test correlations. To control this, we changed the context across sessions. First, a different experimenter administered each session (experimenters were randomly assigned to Sessions 1, 2, and 3 for each child). To ensure consistency across experimenters, a senior researcher watched videos of every session. Second, the testing table was rotated and covered with a different color tablecloth to alter the visual context. These changes in the social and perceptual contexts across sessions should reduce between-session contextual priming. To our knowledge, no other study of children's cognition has taken such measures to control spurious shared variance due to priming over repeated testing.

Cognitive flexibility tests

Children completed three verbally cued flexibility tests: 3DCCS, FIM-Ob, and FIM-An. Each test included three blocks of trials defined by different phrase cues or rules. The same stimuli were shown in each block, and across blocks children could switch responses correctly or incorrectly or repeat a prior response. To ensure that first-block responses were accurate, and that children built a response habit in this block, the strongest (i.e., easiest) cue from each test was assigned to the first block, the next-strongest cue to the second block, and the weakest cue to the last block (cue strength was based on data from Deák, 2000; Deák & Narasimham, 2003, 2014; Narasimham et al., 2015). This design holds order-by-difficulty interactions constant across children and tests. It also maximizes the probability that every child has many opportunities to switch responses. This is critical for making measures of flexibility (described below) interpretable. The cue order in the FIM-Ob was "is made of," "has a(n)," and "is shaped like a(n)"; the cue order in the FIM-An was "is a(n)," "holds a(n)," and "lives in/on a(n)." The rule order in the 3DCCS was shape, color, and size.

Stimulus order in each test was randomized for each child and repeated across blocks. Children received nonspecific feedback for every response (i.e., "thank you"). The experimenter maintained eye contact with the child in every trial and used a uniform tone of voice to avoid providing differential feedback.

Three Dimension-Changes Card Sorting

Photoshop-modified clip art images printed on laminated 21-cm^2 cards depicted prototypical familiar animals (dog, fish, and bird) in three focal colors (red, blue, and yellow) and three sizes (approximately 3.3 cm^2, 8.9 cm^2, and 17.2 cm^2). A fourth distracter showed a medium–large green frog. (Stimuli are available at http://cogdevlab.ucsd.edu/resources/.) Distracters were used in each test to check whether children were attentive and compliant. Children sorted five test cards into four white cardboard boxes, each with a different standard on top. For each test card, one standard had the same shape, one the same color, and one the same size, as shown in Fig. 1. Standards differed in all property values, so any match was unambiguous. Each test card had different combinations of properties than any standard, so it would go in a different box under every rule. Test cards were randomized for each child, but any property occurred no more than twice, and no two properties (e.g., small + blue) were combined more than once. Before the test, children were asked to label the animal, color, and size properties of each card to ensure that they knew the relevant labels (e.g., "blue," "fish") and understood the game labels (e.g., "color game"). All children demonstrated comprehension. The rules of the first game were stated three times using different phrasings. Key instructions from the flexibility tests are provided in Appendix B. Before the test trials children were asked to restate the rules and answer several rule comprehension questions (based on Zelazo, 2006). Before the second and third blocks, children were told (three times) to stop playing the old game and start playing a new game. The new rule was explained, and children's comprehension was checked.

Children sorted each of the five cards three times, once per rule (animal, color, or size game). Specific subtask rules (e.g., "dogs go in this box") explicitly indicated where to place each card in a given block.

Flexible Induction of Meaning–Objects

Five sets of novel objects each included a standard and four comparison objects. Each standard matched one of three comparison objects on one of three novel properties: shape, material, or affixed part. The fourth object in each set was a distracter (see Fig. 2 and Deák, 2000). In each of 15 test trials, children were told to look at all of the objects, and then the experimenter said (twice) of the standard either, "This *is made of* [Word 1]," "This *is shaped like a(n)* [Word 2]," or "This *has a(n)* [Word 3]." The experimenter then indicated the comparison objects and asked, "Which one of these also [Cue] [Word]?" The prompt was repeated after 8 s if a child did not answer. Each block featured a different phrase cue. Object positions were randomized on each trial, and words were randomly assigned to properties.

Flexible Induction of Meaning–Animates

The FIM-An test used five sets of five color pictures (12.5 cm^2) of novel creatures (some from Barlowe & Summer, 1979) holding novel objects in novel habitats (see Fig. 3 in Deák & Narasimham, 2014). (Stimuli are available at http://cogdevlab.ucsd.edu.) Each set's standard matched one of three comparison pictures on one of three properties: species, habitat, or held object. The fourth distracter had different properties. In each of 15 trials, children were first told to look at the pictures, and then the experimenter said of the standard either, "This *is a(n)* [Novel Word 1]," "This *lives in/on a(n)* [Word 2]," or "This *holds a(n)* [Word 3]." The experimenter then indicated the comparison pictures and asked, "Which of these also [Cue] [Word]?" Each block featured a different phrase cue. Picture positions and words were randomized as in the FIM-Ob test.

Scoring: flexibility tests

Each flexibility test was evaluated using three measures. First, accuracy was coded as the number of cue- or rule-appropriate responses in each block and in total. Total accuracy indicates sensitivity to cues or rules. In addition, accuracy in the first block across tests provides an index of children's ability to comprehend cues or rules.

Second, flexibility was assessed using a more focused measure that allows comparison across tests, CORSWOPS (correct switches/opportunities; Deák & Narasimham, 2014), or the proportion of correct switches in later blocks corrected for opportunities to switch correctly. This is the proportion of trials in Blocks 2 and 3 when the child chose a cue- or rule-appropriate item that was different from the previous item chosen from that set. Because accuracy in Blocks 1 and 2 can vary, the proportion of *possible* correct switches can differ in Blocks 2 and 3. Correcting for the actual number of opportunities to switch correctly provides an index of flexibility that is less biased by age and other factors.[4] Although CORSWOPS is strongly correlated with total correct responses, it controls for variability in initial accuracy across children. CORSWOPS is a general index that can be compared across any flexibility test that meets a few assumptions: discrete correct and incorrect responses and sufficient post-switch opportunities to derive proportional scores. These assumptions are met by all three tests. For example, even the youngest quartile of our sample (43 months or younger) had enough opportunities to switch (means = 94% of post-switch 3DCCS trials, 79% of FIM-Ob trials, and 99% of FIM-An trials) to derive meaningful CORSWOPS proportions.

Third, children's responses across trials of any flexibility test usually fit some sequential pattern. In previous studies (Deák, 2000; Deák & Narasimham, 2014), children's response patterns could be classified as *flexible* (in the current design with three blocks of 5 trials, this entails 13 or more correct choices with 7 or more correct switches), *partly flexible* (9–12 correct choices with 5 or more correct switches), *perseverative* (7 or fewer correct choices with 3 or fewer switches [correct or not]), or *indiscriminate* (10 or fewer correct choices with 4 or more switches but 3 or fewer correct switches). These categories might reflect different approaches to the test; flexible patterns indicate adaptation to each cue/rule; partly flexible patterns reflect adaptation to two of three cues/rules or inconsistent use of each cue/rule (perhaps with high uncertainty), perseverative patterns reflect either failure to encode cue/rule changes or failure to weight later cues/rules higher than previous responses, and indiscriminate patterns might reflect high uncertainty about mappings of cues/rules to stimulus properties. These patterns, therefore, suggest different sources of error that are not differentiated by parametric measures.

Cognitive tests: executive functions and knowledge

Response speed: Box Completion

Children saw a page with 35 three-sided boxes (Salthouse, 1994), each missing a randomly chosen side. Children were instructed to "close" each box by drawing a line across the open side. After doing a practice row, children completed as many boxes as possible within 60 s. We report the number completed within the first 30 s, which is less affected by conflating variables such as vigilance, distraction, and boredom.

Working memory: Memory For Names

Children heard names for a series of alien creatures and then identified each alien by name (Woodcock & Johnson, 1989). After each trial, a new name was added, so memory load gradually increased. Testing continued until children exceeded a specified number of errors (see Dean & Woodcock, 1999).

Inhibition (lexical): Stroop Day–Night

Children were instructed to say the word "day" when shown a picture of the moon or to say the word "night" when shown a picture of the sun (Diamond, Kirkham, & Amso, 2002; Gerstadt et al., 1994). After completing up to 6 practice trials with feedback, children completed two blocks of 10 trials in quasi-random order without feedback.

Inhibition (action): Tapping

Following Luria (1962/1966; see also Diamond & Taylor, 1996), children were instructed to tap once (with a plastic rod) if the experimenter tapped twice and to tap twice if the experimenter tapped

[4] CORSWOPS does not count trials in which a child repeats a response that was first inappropriate but became appropriate after the cue switch. Even if the second response is "correct," it is not counted as a correct *switch* because there was no *opportunity* to switch correctly and no way to distinguish flexible responding from perseverative responding. It is most conservative to exclude these responses entirely. Fortunately, these cases are rare (mean = 5.4% of post-switch responses overall) and do not affect the results.

once. After practice trials with feedback, children completed two blocks of 10 trials in quasi-random order without feedback.

Lexical knowledge: receptive vocabulary

The PPVT-3 was administered according to standard procedures (Dunn & Dunn, 1997).

Conceptual knowledge: causal inference

Based on Das Gupta and Bryant (1989), we showed children photographs (16 cm^2) that implied events in which objects underwent changes (e.g., broken flowerpot and glued-together flowerpot). The experimenter described the pictures in general terms ("Look at this. First it looked like this. ... Then I did something to it, and now it looks like this."). Children then were shown photographs of four possible instruments (e.g., hammer, light bulb, brush, glue) and were asked to choose the one that caused the change ("Which of these things ... [made] it like this?"). Children completed 2 practice trials with feedback and then eight test problems without feedback. The latter included four easier problems and four harder ones (see Das Gupta & Bryant, 1989), to increase variability. Items are described in Appendix C. Item order and picture position were randomized for every child.

Scoring

All responses were coded online by a second researcher. Videotapes were recoded for accuracy by an independent researcher. Online accuracy was greater than 98%. Box Completion scores were the number of boxes "closed" within 30 s. Standard scoring rules were used for Memory for Names and PPVT-3. Total correct was calculated for Stroop Day–Night, Tapping, and causal reasoning.

Results

The main and replication groups performed nearly identically on all tests (all *t*s < 1.5), so they were combined in all further analyses ($N = 95$). There were no gender effects on any task (all *t*s < 1), so girls and boys were combined.

Flexibility task performance

Cue/rule accuracy is shown in Fig. 4. Mean accuracy in each test was higher in Block 1 than in Blocks 2 and 3. This could be due to limitations in flexibility, increasing difficulty of later cues/rules, or both. All three tests show a negative quadratic trend across blocks in repeated-measures analyses of variance (ANOVAs) with Greenhouse–Geisser correction. The within-participants effect was significant in the 3DCCS, $F(1.3, 124) = 38.4$, $p < .001$, $\eta^2 = .29$, in the FIM-Ob, $F(1.9, 179) = 16.8$, $p < .001$, $\eta^2 = .15$, and in the FIM-An, $F(1.7, 159) = 92.6$, $p < .001$, $\eta^2 = .50$.

Although the tests were designed to have similar difficulty, FIM-Ob accuracy was lower (mean correct = 57.4%, $SD = 22.6$) than FIM-An accuracy ($M = 67.3\%$, $SD = 27.2\%$) or 3DCCS accuracy ($M = 68.4\%$, $SD = 29.6\%$), one-way ANOVA, $F(2, 188) = 10.6$, $p < .001$. However, this does not present a major interpretive problem because the mean difference is only 11%, and variance is similar across tests, with no ceiling or floor effects.

Flexibility was similar across tests; mean CORSWOPS was 54.6% in the 3DCCS ($SD = 42.3\%$), 47.1% in the FIM-Ob ($SD = 32.7\%$), and 53.2% in the FIM-An ($SD = 39.5\%$), $F(2, 188) = 1.9$, $p = .148$, *ns*, one-way ANOVA. Correlations between age and flexibility (CORSWOPS) also were similar across tests ($r = .498$ in FIM-Ob, $r = .532$ in FIM-An, and $r = .500$ in 3DCCS, all *p*s < .001). Fig. 5 shows the distribution of CORSWOPS by age in each test. No test shows a nonlinear inflection or bimodal distribution that would indicate a qualitative shift from 3 to 4 years of age.

Individual differences in flexibility are also evident in qualitative response patterns. The number of children producing each of four patterns in each test, and in each pair of tests, is shown in Appendix D. The distribution differs across tests, $\chi^2(df = 6, N = 95) = 22.5$, $p < .001$. The distribution was similar in the FIM-An and 3DCCS, where 42 (or 45%) of children were flexible, 13 (14%) were partly flexible, 25

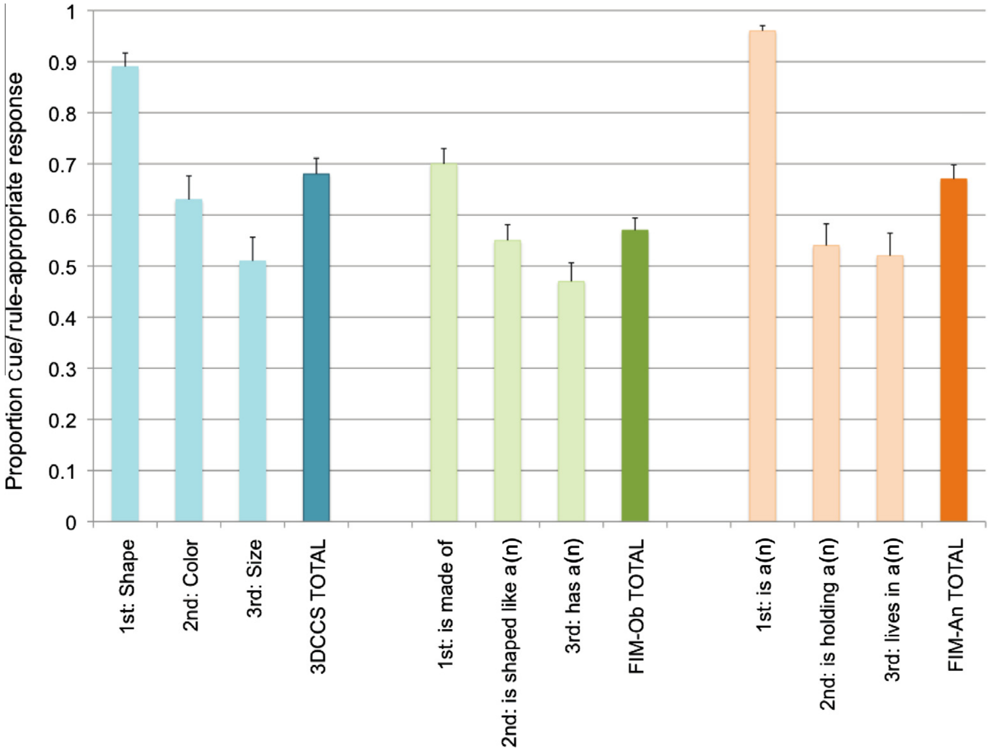

Fig. 4. Mean appropriate responses to each rule or cue in the 3DCCS, FIM-Ob, and FIM-An tests and total appropriate responses. A decline in later blocks is apparent. Error bars represent standard errors.

(29%) were perseverative, and 13 (19%) were indiscriminate. By contrast, in the FIM-Ob only 22% were flexible, 54% were partly flexible, and 37% were indiscriminate, confirming that this test was harder. Despite this, as described below, children tended to produce the same response pattern in both FIM tests.

Coherence between flexibility tasks

Partial correlations, with age removed, between CORSWOPS on the three flexibility tests and EF and language/knowledge test scores, are shown in Table 1. Critical levels were set at $\alpha = .01$ to reduce test-wise Type I error rate. FIM-Ob and FIM-An were strongly related ($r_{Part} = .61$, $p < .001$, $R^2 = .37$). FIM-Ob and 3DCCS were reliably but weakly correlated ($r_{Part} = .27$, $p = .009$, $R^2 = .07$). FIM-An and 3DCCS were not significantly related ($r_{Part} = .12$, $p = .255$).

To verify the robustness of these results, we explored partial correlations using alternate measures, including number of correct switches in flexibility tests uncorrected for number of opportunities, *z*-scores instead of totals for the inhibition tests, and raw PPVT-3 scores. In all cases, the coefficients were nearly identical and retained the same level of statistical significance.

We also tested whether verbal knowledge (e.g., cue/rule comprehension) could explain the strong association between FIM tests. Partial correlations among flexibility tests (CORSWOPS) were calculated, with age, vocabulary, and correct Block 1 responses on all three tests (indicating cue/rule comprehension) partialled out. The correlation between FIM tests remained strong ($r_{Part} = .64$, $p < .001$, $R^2 = .41$). The relation between FIM-Ob and 3DCCS remained reliable but weak ($r_{Part} = .28$, $p = .007$,

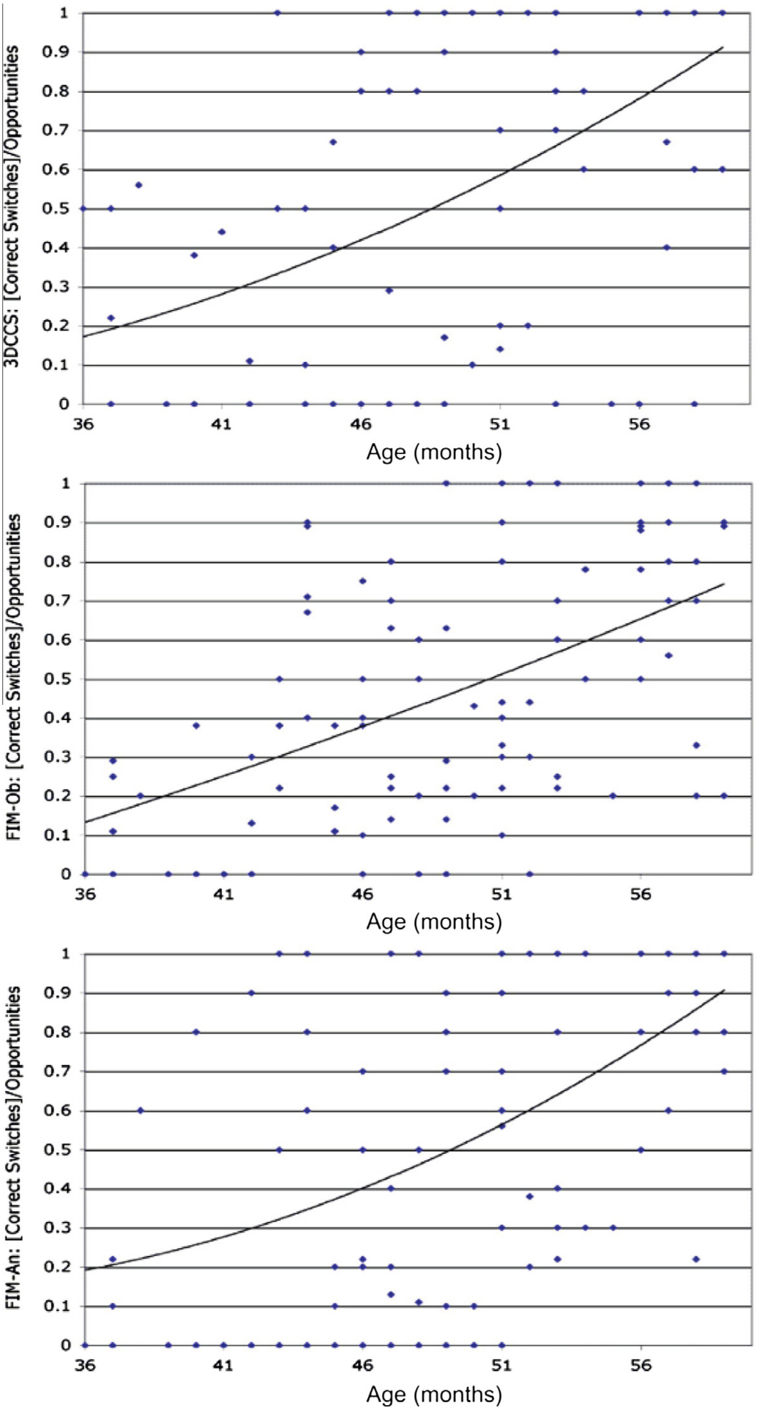

Fig. 5. Scatterplots of CORSWOPS in three flexibility tests, with regression lines: FIM-Ob (top), FIM-An (middle), and 3DCCS (bottom). Scores are arranged by age. The best-fitting trend for each test is nearly linear. No discontinuity between 3- and 4-year-olds (e.g., perseverative vs. flexible) is evident, contrary to a possible interpretation of results from binary rule-switching tests (e.g., DCCS).

Table 1
Partial correlations, controlling for age, among flexibility tests (bold outline), EF measures (Box Completion [speed], Stroop Day–Night [verbal inhibition], Tapping [action inhibition], and Memory for Names [working memory]), and knowledge tests (PPVT-3 and causal inference).

	2	3	4	5	6	7	8	Causal inference
1. FIM-Ob	.61***	.27*	.31**	.28*	.30**	.13	.30**	.31**
2. FIM-An		.12	.30**	.19	.23	.04	.35***	.27*
3. 3DCCS			.26	.17	.24	.06	.31**	.41***
4. Boxes				.16	.07	.08	.07	.21
5. Stroop Day–Night					.43***	.32**	.21	.05
6. Tapping						.33**	.51***	.26
7. Memories for Names							.32**	–.02
8. PPVT-3								.38***

Note: Dependent measures (FIM-Ob, FIM-An, and 3DCCS): CORSWOPS; Boxes: boxes completed in 30 s; Stroop Day–Night and Tapping: total correct; Memory for Names and PPVT-3: standardized scores; causal inference: total correct.
* $p < .01$.
** $p < .005$.
*** $p < .001$.

$R^2 = .08$). The relation between FIM-An and 3DCCS remained nonsignificant ($r_{Part} = .08$, *ns*). These results indicate that cue comprehension cannot fully explain consistency between FIM tests.

We also examined consistency in individual children's response patterns (flexible, partly flexible, perseverative, and indiscriminate) between tests. Appendix D shows the number of children who produced the same patterns on each pair of tests. On the FIM-Ob and FIM-An tests, 48.4% of children ($n = 46$) produced the same pattern—twice the percentage expected (24.7%) based on marginal cross-products and nearly identical to the proportion (50%) reported by Deák and Narasimham (2014). By contrast, only 30.5% of children ($n = 29$) produced the same pattern on the FIM-Ob and 3DCCS, just slightly above the expected number (23.9%). Finally, 38.9% ($n = 37$) produced the same pattern on the FIM-An and 3DCCS, also just slightly above the expected number (30.6%). Thus, only the FIM tests yielded more concordant response patterns than expected.

Executive function and reasoning tests

Results from EF tests and language/conceptual knowledge tests are shown in Table 2. Age was significantly related to response speed, lexical and action inhibition, vWM, and causal inference.

Table 2
Children's performance on EF and language/conceptual knowledge tests.

Test	Mean score (SD)	Correlation with age	FIM-Ob predictor β	FIM-An predictor β	3DCCS predictor β
Response speed (Box Completion)	13.6 (4.8)	$r = .55$, $p < .001$	.22	.29	.18
Inhibition (verbal) (Stroop Day–Night)	13.4 (5.0)	$r = .29$, $p = .005$	.29		
Inhibition (action) (Luria Tapping)	13.0 (5.7)	$r = .63$, $p < .001$	.19		.25
Verbal working memory (Memory for Names)	34.3 (14.1)	$r = .33$, $p < .001$			
PPVT-3 (lexical knowledge)	109.8 (12.5)	$r = .16$, $p = .122$	.18	.53	
Causal inference (Das Gupta & Bryant, 1989)	4.94 (1.86)	$r = .55$, $p < .001$	.50		.56

Note: Pearson's correlations with age are shown with p values. The last columns summarize regressions (see text). Scores: boxes completed (in 30 s), inhibition tests (total correct), Memory for Names (calculated score), PPVT-3 (standardized score), and causal inference (total correct). Regression summary columns show marginal or significant adjusted β weights (all others ns).

Standardized PPVT-3 scores indicate that the sample had somewhat higher lexical knowledge than norming samples.

To explore how these measures related to flexibility, we ran stepwise regressions on CORSWOPS in each test, entering age and the main dependent measure from each EF or knowledge test. The criterion for entry was set at $\alpha = .05$.

For FIM-Ob, three factors significantly and uniquely predicted flexibility: causal inference (Step 1, $\beta_{Std} = .50$, $t = 5.3$, $R^2 = .249$, $p < .001$), verbal inhibition/Stroop Day–Night (Step 2, $\beta_{Std} = .295$, $t = 3.2$, $R^2_{change} = .083$, $p = .002$), and response speed/boxes (Step 3, $\beta_{Std} = .22$, $t = 2.2$, $p = .031$, $R^2_{change} = .037$, $p = .031$). The three-factor model accounted for $R^2_{Adj} = .346$ ($SE = .26$), $F(1,83) = 16.2$, $p < .001$. Two other factors were marginally significant: vocabulary/PPVT ($\beta = .182$, $t = 1.7$, $p = .086$) and action inhibition/Tapping ($\beta = .195$, $t = 1.7$, $p = .089$).

For FIM-An, two factors significantly and uniquely predicted flexibility: vocabulary (Step 1, $\beta_{Std} = .53$, $t = 5.7$, $R^2_{Adj} = .271$, $p < .001$) and response speed (Step 2, $\beta_{Std} = .29$, $t = 3.1$, $R^2_{change} = .075$, $p < .001$). The two-factor model accounted for $R^2_{Adj} = .339$ ($SE = .319$), $F(1,84) = 23.1$, $p < .001$. Another factor, age, was marginally significant, $\beta = .215$, $t = 1.9$, $p = .057$.

For 3DCCS, two factors predicted flexibility: causal inference (Step 1, $\beta_{Std} = .56$, $t = 6.2$, $R^2_{Adj} = .304$, $p < .001$) and action inhibition/Tapping (Step 2, $\beta_{Std} = .25$, $t = 2.5$, $R^2_{change} = .047$, $p = .015$). These accounted for $R^2_{Adj} = .344$ ($SE = .341$), $F(1,84) = 23.5$, $p < .001$. Another factor, response speed/boxes, was marginally significant ($\beta = .184$, $t = 1.9$, $p = .066$).

The regression results suggest that some common predictor abilities might explain the correlation between FIM tests. To assess this, we calculated partial correlations among flexibility tests (CORSWOPS), removing not only age but also all factors that predicted significant variance in multiple flexibility tests: response speed, action inhibition, causal reasoning, and vocabulary. With all of these factors partialled out, the strong association between FIM tests remained ($r_{Part} = .63$, $p < .001$, $R^2 = .40$), as did the weak partial correlation between the FIM-Ob and 3DCCS tests ($r_{Part} = .22$, $p = .037$, $R^2 = .05$). The former is significantly stronger than the latter ($z = 3.5$ by Fisher transformation, $p < .001$). The correlation between FIM-An and 3DCCS remained nonsignificant ($r_{Part} = .03$).

Discussion

This study compared English-speaking preschool children's performance on three flexibility tests in relation to executive functions and verbal and conceptual knowledge. There was a strong correlation between two tests of flexible induction of word meanings, independent of variance due to age, response speed, inhibition, or verbal knowledge. Flexibility in one cue induction test, the FIM-Ob, was also correlated with flexibility in the rule-switching 3DCCS test. However, this correlation was significantly weaker ($R^2 = .05$). In addition, children tended to produce the same response patterns on both FIM tests, but were no more likely than chance to produce the same pattern on an FIM test and the 3DCCS. Thus, parametric and categorical measures both suggest that individual cue-induction flexibility for word meanings was highly stable across tests, but was at best weakly related to flexibility of rule switching. This was true even though the FIM-An was more similar to the 3DCCS in overall difficulty and in some methodological factors (e.g., stimulus material, stimulus domain).

One interpretation is that some underlying capacity contributes to high individual stability of cue-induction flexibility for meaning interpretation, but not to rule-switching flexibility. We cannot say how general the capacity is; it might pertain just to word meanings, or to broader semantic inferences, or perhaps to a wide variety of probabilistic cues (e.g., nonverbal behaviors). Regardless, the results show that it is inappropriate to treat a single test of flexibility in children as measuring some general capacity. This confirms other recent evidence; Ramscar and colleagues (2013) showed that at least two processes contribute to children's rule-switching flexibility, explaining why different DCCS versions yield different results (e.g., Perner & Lang, 2002). In addition, adult studies suggest that flexibility is task dependent (e.g., Kim et al., 2011; Luwel et al., 2003; Ravizza & Carter, 2008).

Could the correlation between FIM-An and FIM-Ob flexibility be due to shared method variance instead of cue-induction flexibility? We cannot rule this out entirely, but it is noteworthy that all stimulus items, the stimulus medium (pictures vs. objects), key stimulus properties, verbal cues, the content domain, and the words themselves were all entirely different between tests. In addition, the tests were administered on different days by different experimenters in a different visual environment. Thus, many aspects of the test were changed. However, some aspects were similar across FIM tests: on every trial, the experimenter presented a novel multidimensional stimulus, told the child a new fact about it (including a novel word), and asked the child to generalize the word to one of four other stimuli. These methodological similarities might have contributed to the between-test correlation. It will require further investigation to completely separate the causes of shared variance. However, even if some portion of shared variance is due to shared methods variance, that would strengthen one conclusion from this study: that cognitive flexibility cannot be considered a global executive capacity in children and should not be estimated from a single test measure. After all, when intercorrelated factors are partialled out, even the FIM tests share less than 40% of variance.

Relations to executive functions

The results also address how several EFs (processing speed, inhibition, and vWM) relate to children's flexibility. If any EF had a consistent relation with all flexibility tests, it would suggest a stable underlying factor or contributor to task-general flexibility processes (Miyake et al., 2000). The results are equivocal in this regard: on one hand, regression analyses showed a different subset of predictors of each flexibility test, failing to confirm a general processing model. On the other hand, response speed was a significant or marginal unique predictor of flexibility in all three tests. This supports the view that processing speed is a general predictor of higher order cognitive and linguistic processes and fluid intelligence (e.g., Kail & Hall, 1994; Li et al., 2004; Salthouse, Fristoe, McGuthry, & Hambrick, 1998). Notably, it is also a predictor of older children's cognitive flexibility, at least in rule-switching tests (Cepeda et al., 2000). A recent study (Holt & Deák, 2015) extended this finding to preschool-aged children. However, in the current study, response speed in a visuomotor test (Box Completion) only marginally predicted rule-switching flexibility. Similarly, Cepeda and Munakata (2007) did not find that 5- and 6-year-olds' speed uniquely predicted flexibility in a computerized 3DCCS test. Thus, even the relation between response speed and rule switching is not consistent across studies. Because studies have used different measures of speed and flexibility as well as different ages, it is currently impossible to determine why this is so. That would require a study with multiple measures of both factors with varied task demands.

Cognitive inhibition has been hypothesized to contribute to cognitive flexibility. However, the current results suggest that children's flexibility is not restricted by their ability to inhibit verbal associations or responses. The Stroop Day–Night test, which requires inhibiting and reversing verbal associations (Simpson & Riggs, 2005), predicted 8% of variance in FIM-Ob and did not predict unique FIM-An or 3DCCS variance. Deák and Narasimham (2003, 2014) found no relation between the Stroop Day–Night test and the FIM-Ob or FIM-An test. Because the current finding of a weak but reliable correlation between the Stroop Day–Night and FIM-Ob tests is inconsistent with those previous results, it might indicate a context-specific association, or sampling or Type I error. Regardless, the sum of available evidence does not suggest that verbal inhibition is a limiting factor in young children's flexibility.

Luria's Tapping test, which requires children to inhibit action imitation, was a reliable but minor predictor (R^2 = .05) of 3DCCS flexibility and a marginal predictor of FIM-Ob flexibility. In spite of this, there are limitations to inhibition-based accounts of cognitive flexibility. One is that cognitive inhibition itself might not be a coherent trait. Although the Stroop Day–Night and Tapping tests were correlated in the current data (see also Montgomery & Koeltzow, 2010), they shared only 18% of variance, suggesting mostly non-shared rather than shared processes. This confirms other evidence that children's performance varies considerably across inhibition tests (Carlson et al., 2002). One possible explanation is that there are multiple inhibitory processes that are all elicited to varying degrees by different tests of inhibition and (less directly) by different tests of flexibility, such that the association between any two tests cannot currently be predicted. This hypothesis has not been explored, but it is consistent with existing evidence (Blackwell, Chatham, Wiseheart, & Munakata, 2014; Cepeda et al., 2000; Holt & Deák, 2015). In addition, although Zelazo and colleagues (2003) claimed that negative priming, an inhibitory process, affects preschoolers' rule switching, Ramscar and colleagues (2013) showed that the relevant findings can be explained by associative learning processes. In sum, previous and current results do not point to a clear specific causal relation between developing inhibitory mechanisms and children's cognitive flexibility.

Recent evidence has suggested a relation between vWM efficiency and cognitive flexibility in children as well as adults (Cepeda & Munakata, 2007; Gruber & Goschke, 2004; Holt & Deák, 2015; Schneider & Logan, 2009). However, we found no relation between Memory for Names performance and any flexibility measure. This is notable because the FIM tests could require vWM for novel words, and the 3DCCS requires vWM for the current rule. Yet these negative results converge with prior findings that children's vWM capacity does not predict their flexibility in verbally cued tests (Deák & Narasimham, 2003; Zelazo et al., 2003). One possible reason for these negative findings is that vWM *capacity* is dissociated from vWM *efficiency* or *specificity* of retrieval and/or maintenance (e.g., Postle, Berger, & D'Esposito, 1999). It seems that children's cognitive flexibility is unrelated to variability in vWM capacity, but perhaps it is still somehow related to vWM efficiency or specificity. A question for future study is whether cue-induction flexibility and rule-switching flexibility are equally sensitive to differences in vWM processes related to updating, maintenance, or retrieval.

Relations to knowledge

Cognitive flexibility is critical for everyday language use (Deák, 2003) and requires semantic knowledge, at least when task cues are linguistic (see Hermer-Vazquez et al., 2001). The flexibility tests used here required comprehension of, and response to, verbal cues. Thus, any correlations between flexibility tests could have been due to receptive language knowledge. To test this, receptive vocabulary was assessed (Sattler, 2002). PPVT-3 vocabulary predicted 27% of variance in FIM-An flexibility and marginally predicted FIM-Ob flexibility. Deák and Narasimham (2003) also found a marginally significant relation of vocabulary with FIM-Ob flexibility, but a nonsignificant (positive) correlation with FIM-An flexibility. Thus, there is some converging evidence that individual differences in receptive vocabulary predict children's word-induction flexibility. However, even with vocabulary partialled out—along with accuracy in the first blocks of flexibility tests—a strong correlation remained between FIM tests, suggesting that verbal knowledge did not mediate the association. This is not too surprising, as cues and rules were chosen to be comprehensible to typical 3-year-olds. Still, differences in children's certainty or speed of cue/rule processing might have affected their ability to use cues flexibly. However, the results do not support this hypothesis. Another interpretation of the correlation between vocabulary and FIM scores is that word-learning flexibility makes a small but cumulative contribution to children's vocabulary. That is, children who can more flexibly select changeable, probabilistic contextual cues to infer novel word meanings might acquire new words faster, all else being equal, than less flexible children.

It is also possible that conceptual knowledge contributes to cognitive flexibility. This hypothesis has received little attention (but see Bilalic, McLeod, & Gobet, 2008). A test of causal inferences of object effects (Das Gupta & Bryant, 1989), with no flexibility demands and minimal verbal demands, uniquely predicted flexibility in the FIM-Ob and 3DCCS. This is not predicted by current accounts (Deák, 2003; Zelazo et al., 2003). How can we explain it?

One possibility is that all three tests share a demand to select one abstract similarity, out of several compelling options, that is most relevant to the given problem, and to ignore at least two conflicting options. The selections cannot rely on habitual responses or repetition, but require trial-by-trial inductive reasoning. Children might vary in this capacity, which is consistent with a description of "fluid intelligence" by Horn and Cattell (1966): "perceiving relations, educing correlates, maintaining ... awareness in reasoning, and abstracting ... figural, symbolic, and semantic content" (p. 268). Although this explanation is descriptive rather than explanatory, it points to other relevant efforts to elucidate the relation between concepts of fluid intelligence and executive functions (e.g., Decker, Hill, & Dean, 2007; Kane & Engle, 2002). These efforts have yet to extend to research on the development of cognitive flexibility, but the current result suggests that this might be a fruitful research direction. However, note that the causal inference test was not significantly correlated with the FIM-An test, so the finding might not be very general. In addition, when causal inference scores were partialled out, the association between the FIM tests remained strong, so it cannot fully explain the between-test coherence.

General implications

An interpretation consistent with all available data is that children's flexibility is determined by multiple factors, including (a) processing factors related to task type, for example, cue induction for inferences of meaning, or rule switching; (b) subtask-specific factors (e.g., understanding specific cues or rules; Munakata & Yerys, 2001); (c) cognitive moderators, including response speed (Cepeda, Blackwell, & Munakata, 2013; Cepeda et al., 2001), working memory efficiency (Cepeda & Munakata, 2007; Holt & Deák, 2015), and possibly (d) a faculty for selecting abstract relations for novel inferences (i.e., "fluid intelligence"). These factors together predict considerable variability in children's flexibility.

The current results also confirm that children should not be classified simply as "flexible" or "perseverative." That dichotomy is an artifact of low-sensitivity test paradigms (e.g., DCCS). Young children, like adults, produce distinct perseverative and indiscriminate error patterns. Barceló and Knight (2002) speculated that adult frontal patients' indiscriminate errors are related to working memory inefficiency, but it remains to be determined whether children produce indiscriminate response patterns due to inefficient vWM. However, most children (80%) did not produce the same pattern on all three tests, suggesting that children's performance on a given flexibility test cannot be assumed to indicate a generalized deficit or immaturity.

The results also disconfirm an impression from the literature that cognitive flexibility improves qualitatively between 3 and 4 years of age. All three flexibility tests show a positive age-related trend that was nearly linear, with no inflections or discontinuity (Fig. 5). In addition, there was high inter-individual variability at any age stratum. Thus, although age predicts flexibility, it is a poor predictor by itself.

The results leave unanswered questions for future research. One limitation is that this sample was restricted to healthy, English-speaking, middle-class North American children. It is unknown whether the results generalize to other populations. In addition, we could not collect response times, eye movements, or physiological measures (e.g., electroencephalogram, EEG) that might reveal subtler but potentially predictive indicators of age and individual differences. A third limitation is that we used only single measures of executive functions (e.g., response speed, action inhibition). Single measures are nonoptimal because any single test brings idiosyncratic measurement error; a latent variables approach is preferable. The current results, therefore, provide suggestions for future investigations rather than generalizable estimates of the associations among latent cognitive factors. Another limitation is that in all three flexibility tests the cue/rule order got progressively harder. Although this made between-test individual differences interpretable, it introduces the possibility that the results will not generalize to other subtask orders (e.g., hard-to-easy test situations). However, Deák and Narasimham (2014) also found a strong correlation between the FIM-An and FIM-Ob tests without this constraint. Nonetheless, order-specific between-test correlations should be evaluated in future studies. Finally, in future studies it would be ideal to obtain independent estimates of each child's comprehension of each cue or rule. Although adequate cue-comprehension estimates are almost never obtained in studies of children's EF or cognitive flexibility, they provide important information (Munakata & Yerys, 2001). Fortunately, the correlation between FIM tests persisted when Block 1 accuracy (an index of cue comprehension) and vocabulary were partialled out, indicating that these were not determining factors. This result, therefore, confirms very limited cross-test consistency of flexibility in young children, particularly in cue-based induction of word meanings.

Acknowledgments

This work was supported by grants from the National Science Foundation (NSF-BCS0902027) and from the UCSD Academic Senate to G. Deák. Thanks go to Elaine Blank, Sean Marco, Ali Moeller, Sam Sedlik, Mieke VanderBourght, Cherry Vu, and Rachel Weisser for assistance in data collection and coding. Thanks also go to Gabriel Catalano and Annelise d'Souza for helpful comments on earlier drafts. Special thanks go to all of the children who participated.

Appendix A.

Test order in main sample and replication sample

Main sample	Session 1	Session 2	Session 3
	[Matching game]	Luria Tapping	Box Completion
	3DCCS	FIM-Ob	Stroop Day–Night
	[Flexible Tool-Use test]	PPVT-3	FIM-An
	ID (Contents 1)	ID (Words)	ID (Object 1, then Color)
		Memory for Names	Causal reasoning
		ID (Contents 2)	ID (Object 2)
Replication sample	Stroop Day–Night	Memory for Names	Luria Tapping
	PPVT-3	[Flexible Tool-Use test]	FIM-Ob
	Box Completion	ID (Object 2)	[Matching game]
	ID (Color)	Causal reasoning	ID (Words)
	FIM-An		3DCCS
	ID (Object 1)		ID (Contents 2)

Note: The number of tests is not matched across days because the tests varied widely in duration (e.g., from 1–3 min for ID test to >20 min for FIM-Ob test). ID, Indeterminacy detection.

Appendix B.

Key instructions in the cognitive flexibility tests

Test	Instruction
FIM-Ob initial instruction	"First look at this one [E is holding standard up for child to see]. Let me tell you something about it. This one is made of [Word 1]. See, it's made of [Word 1]. Now look at these [second experimenter hands first experimenter the first box of objects with the lid already removed]. Can you find me another one that is made of [Word 1] just like this one? [E holds up standard when she says "just like this one" and keeps it held above the comparison objects]."
Switch instruction (example)	"Now I am going to show you these things again, but I am going to tell you something new about them. Remember this one? [touching standard]. Let me tell you something new about it. This one is shaped like a(n) [Word 2]. See, it's shaped like a(n) [Word 2]. Now look at these [presenting comparison objects]. Can you find me another one that is shaped like a(n) [Word 2] just like this one?"

(*continued on next page*)

Appendix B (*continued*)

Test	Instruction
FIM-An initial instruction	"Now we are going to play a game with some cool pictures of space creatures. I am going to show you some pictures, and then I am going to tell you something about them. Let's try it! [placing standard picture]. See this one? This one is a(n) [Word 1]. See, he is a(n) [Word 1]. Now let me show you some more pictures [placing comparison pictures]. Can you find me another one that is a [Word 1] just like this one?"
Switch instruction (example)	"Now I am going to tell you something new about these creatures. Are you ready? Look at this one [placing standard]. Remember this one? This one is holding a(n) [Word 2]. See, he is holding a(n) [Word 2]. Now look at these [placing comparison pictures]. Can you find me another one that is holding a(n) [Word 2] just like this one?"
3DCCS initial instruction	"Now we are going to play the animal game. Let me tell you how to play the animal game. In the animal game, all dogs go in here, all fish go in here, and all birds go in here [pointing]. So, do you see this picture of a dog here? That's to remind you that all dogs go in here. And do you see this picture of a fish ... [etc.]? So, all dogs go in here, all fish go in here, and all birds go in here. Are you ready to play the animal game?"
Switch instruction (example)	"Are you ready to play a new game? We're going to play the color game. Let me tell you how to play the color game. In the color game, all blue things go in here, all red things go in here, and ... [pointing]. So, do you see this blue thing here? That's to remind you that all blue things go in here. Do you see this red thing ... [etc.]? So, all blue things go in here, all red things go in here, and all yellow things go in here. Are you ready to play the color game?"

Appendix C.

Items in causal inference test

Pre → Post event photographs	Choices
Practice problems	
tomato → sliced tomato	WHISK, SPATULA, MEASURING CUPS, KNIFE
torn shirt → sewn shirt	TEAPOT, ROLLERSKATE, MUG, NEEDLE/THREAD
Test problems: Easier	
spilled dirt → swept dirt	CHAIR, TISSUE, CLOCK, BROOM
raw egg → cooked egg	BLENDER, DRYING RACK, NAPKIN HOLDER, STOVE
messy hair → brushed hair	SPONGE, TOOTHBRUSH, ROLLING PIN, HAIRBRUSH
torn paper → taped paper	KEYS, TOY BLOCKS, CRAYONS, TAPE
Test problems: Harder	
wet plate → dry plate	SINK, MICROWAVE, CALCULATOR, TOWEL
chalkboard with writing → erased board	CHALK, SCISSORS, STAPLER, ERASER
broken flowerpot → fixed pot	HAMMER, LIGHT BULB, PAINT BRUSH, GLUE
dirty shirt → clean shirt	KETCHUP, PURSE, IRON, DETERGENT

Note. Based on Das Gupta and Bryant (1989).

Appendix D.

Cross tabulation of individual response patterns in the FIM-Ob, FIM-An, and 3DCCS

FIM-Ob \ FIM-An	Flexible	Partly flexible	Indiscriminate	Perseverative	Total
Flexible	**19**	2	0	0	21
Partly flexible	12	**1**	1	2	16
Indiscriminate	9	6	**12**	8	35
Perseverative	0	4	5	**14**	23
Total	40	13	18	24	[**46 of**] 95

FIM-Ob \ 3DCCS	Flexible	Partly flexible	Indiscriminate	Perseverative	Total
Flexible	**15**	4	0	2	21
Partly flexible	12	**2**	1	1	16
Indiscriminate	8	4	**5**	18	35
Perseverative	8	2	6	**7**	23
Total	43	12	12	28	[**29 of**] 95

3DCCS \ FIM-An	Flexible	Partly flexible	Indiscriminate	Perseverative	Total
Flexible	**23**	6	5	9	43
Partly flexible	8	**3**	0	1	12
Indiscriminate	2	1	**3**	6	12
Perseverative	7	3	10	**8**	28
Total	40	13	18	24	[**37 of**] 95

Note: Cells on concordant diagonal, indicating numbers of children who produced the same pattern on both tasks, are indicated in bold.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288–318.
Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130*, 641–657.
Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia, 40*, 349–356.
Barkley, R. (2012). *Executive functions: What they are, how they work, and why they evolved*. New York: Guilford.
Barlowe, W., & Summer, I. (1979). *Barlowe's guide to extraterrestrials*. New York: Workman.
Bilalić, M., McLeod, P., & Gobet, F. (2008). Inflexibility of experts—reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology, 56*, 73–102.
Blackwell, K. A., Chatham, C. H., Wiseheart, M., & Munakata, Y. (2014). A developmental window into trade-offs in executive function: The case of task switching versus response inhibition in 6-year-olds. *Neuropsychologia, 62*, 356–364.
Blaye, A., Bernard-Peyron, V., Paour, J. L., & Bonthoux, F. (2006). Categorical flexibility in children: Distinguishing response flexibility from conceptual flexibility—The protracted development of taxonomic representations. *European Journal of Developmental Psychology, 3*, 163–188.
Blaye, A., & Bonthoux, F. (2001). Thematic and taxonomic relations in preschoolers: The development of flexibility in categorization choices. *British Journal of Developmental Psychology, 19*, 395–412.

Bohlmann, N. L., & Fenson, L. (2005). The effects of feedback on perseverative errors in preschool aged children. *Journal of Cognition and Development, 6*, 119–131.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M., Dawson, D. R., et al (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*, 194–209.

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

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Cartwright, K. B., Marshall, T. R., Dandy, K. L., & Isaac, M. C. (2010). The development of graphophonological–semantic cognitive flexibility and its contribution to reading comprehension in beginning readers. *Journal of Cognition and Development, 11*, 61–85.

Centers for Disease Control. (2013, December). Births: Final data for 2012. *National Vital Statistics Reports*. Available online: <http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_09.pdf>.

Cepeda, N. J., Blackwell, K. A., & Munakata, Y. (2013). Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control. *Developmental Science, 16*, 269–286.

Cepeda, N. J., Cepeda, M. L., & Kramer, A. F. (2000). Task switching and attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology, 28*, 213–226.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. (2001). Changes in executive control across the life-span: Examination of task switching performance. *Developmental Psychology, 37*, 715–730.

Cepeda, N. J., & Munakata, Y. (2007). Why do children perseverate when they seem to know better: Graded working memory or directed inhibition? *Psychonomic Bulletin and Review, 14*, 1058–1065.

Chen, Z. (1999). Schema induction in children's analogical problem solving. *Journal of Educational Psychology, 91*, 703–715.

Das Gupta, P., & Bryant, P. E. (1989). Young children's causal inferences. *Child Development, 60*, 1138–1146.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.

Deák, G. O. (2000). The growth of flexible problem solving: Preschool children use changing verbal cues to infer multiple word meanings. *Journal of Cognition and Development, 1*, 157–191.

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. In R. Kail (Ed.). *Advances in child development and behavior* (Vol. 31, pp. 271–327). San Diego: Academic Press.

Deák, G. O., & Boddupalli, G. (2015). *Flexible reasoning about object functions by young children*. San Diego: University of California.

Deák, G. O., & Enright, B. (2006). Choose and choose again: Appearance–reality errors and the logic of questioning. *Developmental Science, 9*, 323–333.

Deák, G. O., & Narasimham, G. (2003). Is perseveration caused by inhibition failure? Evidence from preschool inferences about word meanings. *Journal of Experimental Child Psychology, 86*, 194–222.

Deák, G. O., & Narasimham, G. (2014). Young children's flexible use of semantic cues to word meanings: Converging evidence of individual and age differences. *Journal of Child Language, 41*, 511–542.

Deák, G. O., Ray, S. D., & Pick, A. D. (2004). Effects of age, reminders, and task difficulty on young children's rule-switching flexibility. *Cognitive Development, 19*, 385–400.

Dean, R. S., & Woodcock, R. W. (1999). *The WJ-R and Batería-R in neuropsychological assessment* (Research Report No. 3). Itasca, IL: Riverside.

Decker, S. L., Hill, S. K., & Dean, R. S. (2007). Evidence of construct similarity in executive functions and fluid reasoning abilities. *International Journal of Neuroscience, 117*, 735–748.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*, 352–362.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental Psychobiology, 29*, 315–334.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition*. Circle Pines, MN: American Guidance Service.

Ellefson, M. R., Shapiro, L. R., & Chater, N. (2006). Asymmetrical switch costs in children. *Cognitive Development, 21*, 108–130.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation, 44*, 145–200.

Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal–striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia, 31*, 17–28.

Frank, M. J., O'Reilly, R. C., & Curran, T. (2006). When memory fails, intuition reigns: Midazolam enhances implicit inference in humans. *Psychological Science, 17*, 700–707.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201–225.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like Day-Night test. *Cognition, 53*, 129–153.

Gruber, O., & Goschke, T. (2004). Executive control emerging from dynamic interactions between brain systems mediating language, working memory, and attentional processes. *Acta Psychologica, 115*, 105–121.

Hermer-Vazquez, L., Moffet, A., & Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition, 79*, 263–299.

Holt, A. E., & Deák, G. O. (2015). Children's task-switching: Missing our cue? *Journal of Cognition and Development, 16*, 261–285.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253–270.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence, 3*, 105–120.

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology, 30*, 190–200.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review, 17*, 213–233.
Kail, R. (1991). Processing speed declines exponentially during childhood and adolescence. *Developmental Psychology, 27*, 259–266.
Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30*, 949–954.
Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review, 9*, 637–671.
Kim, C., Johnson, N. F., Cilles, S. E., & Gold, B. T. (2011). Common and distinct mechanisms of cognitive flexibility in prefrontal cortex. *Journal of Neuroscience, 31*, 4771–4779.
Klahr, D., & Chen, Z. (2003). Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Development, 74*, 1275–1296.
Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2011). Middle school students' understanding of core algebraic concepts: Equivalence + variable. In J. Cai & E. J. Knuth (Eds.), *Early algebraization* (pp. 259–276). Berlin: Springer.
Kramer, A. F., Cepeda, N. J., & Cepeda, M. L. (2001). Methylphenidate effects on task-switching performance in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1277–1284.
Li, S., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15*, 155–163.
Luria, A. R. (1966). *Higher cortical functions in man* (B. Haigh, Trans.). New York: Basic Books. (Original work published 1962).
Luwel, K., Verschaffel, L., Onghena, P., & De Corte, E. (2003). Strategic aspects of numerosity judgment: The effect of task characteristics. *Experimental Psychology, 50*, 63–75.
Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports, 76*, 623–626.
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*, 49–100.
Montgomery, D. E., & Koeltzow, T. E. (2010). A review of the Day-Night task: The Stroop paradigm and interference control in young children. *Developmental Review, 30*, 308–330.
Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*, 335–337.
Narasimham, G., Deák, G. O., & Wiseheart, M. (2015). *Measuring age and individual variability in rule-switching flexibility: The Three Dimension Changes Card Sorting test (3DCCS)*. Vanderbilt University.
Nguyen, S. P., & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development, 74*, 1783–1806.
O'Reilly, R., & Frank, M. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18*, 283–328.
Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the Dimensional Change Card Sorting task? *Infant and Child Development, 11*, 93–105.
Postle, B. R., Berger, J. S., & D'Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 12959–12964.
Ramscar, M., Dye, M., Gustafson, J. W., & Klein, J. (2013). Dual routes to cognitive flexibility: Learning and response conflict resolution in the Dimensional Change Card Sort task. *Child Development, 84*, 1308–1323.
Ravizza, S. M., & Carter, C. S. (2008). Shifting set about task switching: Behavioral and neural evidence for distinct forms of cognitive flexibility. *Neuropsychologia, 46*, 2924–2935.
Salthouse, T. A. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology, 30*, 240–259.
Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging, 13*, 445–461.
Sattler, J. M. (2002). *Assessment of children* (4th ed.). La Mesa, CA: Sattler.
Schneider, D. W., & Logan, G. D. (2009). Selecting a response in task switching: Testing a model of compound cue retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 122–136.
Simpson, A., & Riggs, K. J. (2005). Factors responsible for performance on the Day-Night task: Response set or semantics? *Developmental Science, 8*, 360–371.
Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *American Journal of Psychology, 104*, 61–87.
Thompson-Schill, S., Ramscar, M., & Evangelia, C. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science, 8*, 259–263.
Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson psychoeducational battery-revised*. Allen, TX: DLM Teaching Resources.
Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301.
Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68* (3, Serial No. 274).
Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science, 7*, 121–126.
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.