DEVELOPMENTAL NEUROPSYCHOLOGY, *20*(3), 573–591

The Flexible Item Selection Task (FIST): A Measure of Executive Function in Preschoolers

Sophie Jacques and Philip David Zelazo
Department of Psychology
University of Toronto

Abstraction and cognitive flexibility were assessed in 197 preschool children at 2, 3, 4, and 5 years of age using the Flexible Item Selection Task, a task adapted from the Visual–Verbal Test (Feldman & Drasgow, 1951). On this new inductive task, children were shown a set of 3 cards and required to select 2 cards that matched each other on 1 dimension (Selection 1) and then to select a different pair of cards that matched each other on another dimension (Selection 2). Thus, 1 of the 3 cards always had to be selected twice according to different dimensions. Two-year-olds failed to understand basic task requirements as assessed by a criterial measure. Three-year-olds did more poorly on Selection 1 than 4- and 5-year-olds (who performed near ceiling), suggesting that 3-year-olds had difficulty with the abstraction component of the task. Four-year-olds did worse than 5-year-olds on Selection 2, suggesting that they had difficulty with the cognitive flexibility component (i.e., difficulty selecting the same card on more than 1 dimension). Results are discussed in terms of the development of executive function.

Researchers have found important changes during the preschool years in several aspects of executive function (see Zelazo, Carter, Reznick, & Frye, 1997, for a review), including rule use (e.g., Diamond & Taylor, 1996; Frye, Zelazo, & Palfai, 1995; Gerstadt, Hong, & Diamond, 1994; Luria, 1959, 1961; Zelazo, Frye, & Rapus, 1996), working memory (e.g., Alp, 1994; Case, Kurland, & Goldberg, 1982; Gordon & Olson, 1998), planning (e.g., Klahr & Robinson, 1981; Welsh, 1991), and error evaluation (e.g., Briars & Siegler, 1984; DeLoache, Sugarman, &

Requests for reprints should be sent to Sophie Jacques, GRIP, Universite de Montreal, C.P. 6128 Succursale Centre-ville, 3050 Edouard Montpetit, Room/Bur. A–126–4, Montreal, Quebec, H3T IJ7 Canada. E-mail: sophie.jacques@umontreal.ca

Brown, 1985; Gelman & Meck, 1983; Jacques, Zelazo, Kirkham, & Semcesen, 1999). However, relatively little is known about the development of abstraction and flexible thinking. In recent years, developmental neuropsychologists have adapted several traditional neuropsychological tests of abstraction and flexible thinking for use with school-age children (e.g., Chelune & Baer, 1986; Chelune & Thompson, 1987; Welsh, Pennington, & Groisser, 1991), but there is a paucity of convenient and suitable neuropsychological tests of these abilities for use with preschoolers.

A classic neuropsychological test of these abilities, the Wisconsin Card Sorting Test (WCST), was developed by Berg and Grant (Berg, 1948; Grant & Berg, 1948) for the specific purpose of assessing in normal adults "[t]he phenomena of 'abstract behavior' and 'shift of set' in thinking" (Berg, 1948, p. 15). The WCST has been an important tool for differentiating among individuals with various types of brain dysfunctions (e.g., Milner & Petrides, 1984). For example, in a well-known study, Milner (1963) found that individuals with lesions to dorsolateral prefrontal cortex were significantly impaired on the WCST relative to individuals with lesions to other cortical areas. Moreover, the WCST has been used to assess neuropsychological functioning in school-age children with a variety of developmental psychopathologies (see Pennington & Ozonoff, 1996, for a review), and normative data on the WCST are also available (e.g., Chelune & Baer, 1986; Chelune & Thompson, 1987; Welsh et al., 1991).

More recently, however, researchers have expressed some dissatisfaction with the WCST (e.g., Delis, Squire, Bihrle, & Massman, 1992; Levine, Stuss, & Milberg, 1995; Pennington & Ozonoff, 1996), in part because failures on the WCST are difficult to interpret given the large number of cognitive processes that need to be intact to perform well on this task. For example, in addition to assessing participants' ability to detect a correct dimension and their ability to switch flexibly between dimensions, successful performance on the WCST depends on the ability to benefit from feedback, the ability to keep the correct dimension in mind over several trials, and the ability to inhibit prepotent responses. Moreover, recent evidence indicates that difficulties on the WCST may not be specific to individuals with lesions to the frontal lobes (see Stuss, Eskes, & Foster, 1994, for a review). Finally, the numerous cognitive and instructional demands of the WCST undermine its utility with preschoolers.

A less well-known neuropsychological test of abstraction and cognitive flexibility is the Visual–Verbal Test developed by Feldman and Drasgow (1951; Drasgow & Feldman, 1957). The original version of the task (Feldman & Drasgow, 1951) involved 43 cards, each of which depicted four objects. On each card, three of the four objects could be grouped according to one dimension, and either the same or a different trio could be grouped according to a second dimension. For example, on one stimulus card, objects consisted of a small white circle, a large white circle, a large black circle, and another large white circle. Participants were asked to group three objects that were alike in some way (e.g., the three large circles) and then to group three objects that were alike in a different way (e.g., the three white circles). Objects could be grouped according to "color, form, size, structural similarities, naming, and function" (Feldman & Drasgow, 1951, p. 56). The first grouping that participants made provided a direct measure of their ability to identify a single dimension (the abstraction component of the task), and the second grouping provided a measure of participants' ability to switch flexibly between dimensions (the cognitive flexibility component), although it also required abstraction. Feldman and Drasgow found that, relative to a comparison group consisting of normal adults, individuals with schizophrenia had particular difficulty detecting a correct second grouping, suggesting that they have difficulty with flexible thinking. This finding has been replicated several times (e.g., Drasgow & Feldman, 1957; Siegel, 1957; Stuss et al., 1983).

As a measure of abstraction and cognitive flexibility, the Visual–Verbal Test possesses several advantages over the WCST. Whereas the WCST requires that participants respond according to a specific dimension across several trials before switching and responding according to another dimension, the Visual–Verbal Test requires that participants select items according to one dimension and immediately switch and select items according to another dimension. For this reason, the Visual–Verbal Test places fewer demands on working memory. In addition, unlike the WCST, performance on the Visual–Verbal Test does not depend on participants' ability to benefit from feedback. Because of the simplicity of the instructions and of the task, poor performance on this task can be more easily interpreted. Furthermore, because of its simplicity, a modified version of the Visual–Verbal Test can even be used with preschoolers.

To assess abstraction and cognitive flexibility in preschoolers, we developed a new task, the Flexible Item Selection Task (FIST), based on the Visual–Verbal Test (Feldman & Drasgow, 1951). Although the FIST is equivalent to the Visual–Verbal Test in essential characteristics, it differs from its predecessor in several respects. For example, it uses child-friendly stimuli, it has fewer trials than the Visual–Verbal Test, it uses fewer and more clearly defined dimensions, and it was developed with tightly controlled counterbalancing procedures (see Method section).

The primary purposes of this study were to determine whether the FIST is appropriate for use with preschoolers as a test of abstraction and cognitive flexibility and whether there are meaningful age-related differences in performance on this task. Children at four ages (i.e., 2-, 3-, 4-, and 5-year-olds) participated in the experiment. To ensure that any differences between age groups were due to differential difficulties with abstraction, cognitive flexibility, or both, and not to possible group differences in understanding and following basic task instructions, criterial trials were also included. To succeed on these criterial trials, children only needed to understand the instructions themselves, which were identical to the instructions used in the FIST; children were not required to abstract or represent flexibly dimensional information on these trials because it was possible for them to use overall perceptual similarity information instead of dimensional information to succeed. In other words, they could perform well even if they relied on a simpler perceptual-matching strategy. If children erred on these trials, they did not receive the task proper.

METHOD

Participants

The sample consisted of sixty 2-year-olds (*M* = 30.64 months, *SD* = 1.64 months, range = 27.4–35.8 months), fifty-three 3-year-olds (*M* = 42.99 months, *SD* = 1.89, range = 38.7–46.6 months), forty-nine 4-year-olds (*M* = 54.70 months, *SD* = 2.00 months, range = 51.1–58.5 months), and thirty-five 5-year-olds (*M* = 66.29 months, *SD* = 2.53 months, range = 60.7–69.6), including 97 girls and 100 boys. The girl:boy ratios were 32:28 for 2-year-olds, 29:24 for 3-year-olds, 24:25 for 4-year-olds, and 12:23 for 5-year-olds. Eight children (six 2-year-olds and two 3-year-olds) were excluded from the analyses because they refused to perform the task in its entirety. Children were recruited from several local day care centers (n = 72) or from a database of children whose parents had expressed an interest in participating in research (n = 125). Informed consent was obtained from all parents of children who participated in the experiment. Children with suspected or known developmental or medical disorders that might affect their performance did not participate in the experiment (one child in a day care center was not tested because he was suspected of having a developmental delay, whereas another was not tested because he was suspected of having a hearing impairment). Likewise, children in day care centers with a poor grasp of the English language were also not tested (n = 3). Demographic information was not collected systematically, although children appeared to come from a variety of ethnic and socioeconomic backgrounds.

Materials and Task Design

Forty-eight 21.5 × 5.5 cm laminated white cards were used. Each card depicted a set of items that were derived from the combination of four dimensions (shape, color, size, and number). Each dimension was represented by one of three cues. The shape dimension was represented by a phone, a pair of socks, and a fish; the color dimension was represented by the colors pink, purple, and orange; the number dimension was represented by one, two, and three items; and the size dimension was represented by small items (the mean rectangular area of each item was approximately 7 cm^2), medium items (approximately 13 cm^2), and large items (approximately 23 cm^2). Thus, a specific card might depict one large pink phone, whereas another card might depict three small purple fish. When only one item was depicted on a card, it was positioned in the center of the card. When two items were depicted, one was positioned at the extreme left of the card and the other at the extreme right. Finally, when three items were depicted, one was located at the extreme left end, another at the center, and the other at the extreme right end of the card.

The FIST consisted of 1 demonstration trial, 2 criterial trials, and 12 test trials.

Demonstration and criterial trials. The demonstration trial and the criterial trials always consisted of sets of four cards. Two of these cards were identical on all four dimensions (i.e., color, shape, size, and number), whereas the other two cards, which were also identical to each other on all four dimensions, differed from the first pair on all dimensions (see Figure 1). For example, for the demonstration trial, two cards depicted one small orange pair of socks and two cards depicted two medium purple fish.

These three trials (demonstration trial and criterial trials) were always presented in the same order across all children. The placement of matching pairs was counterbalanced across the three trials. That is, on the demonstration trial the first card (from top to bottom) matched the fourth card (and therefore, the middle two

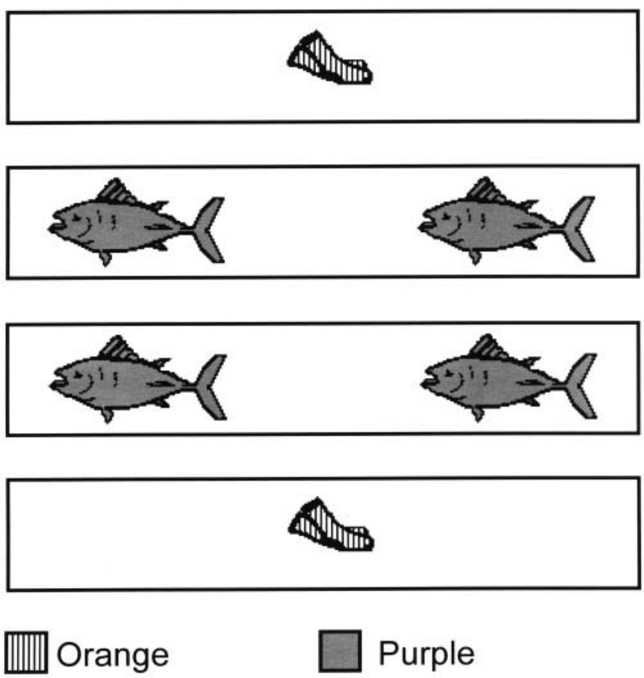

FIGURE 1 Example of cards presented in the demonstration and criterial trials.

cards matched each other), whereas on Criterial Trial 1 the first card matched the second card, and on Criterial Trial 2 the first card matched the third card. Additionally, each cue of each dimension (e.g., orange for color or small for size) was used twice across these three trials.

Test trials. Test trials each consisted of sets of three cards that were identical on two (irrelevant) dimensions (e.g., size and number) but differed on two relevant dimensions (e.g., color and shape). Two of the three cards matched on one of the relevant dimensions (e.g., shape), and a different pair of cards matched on the other relevant dimension (e.g., color). For example, one card might depict one medium purple phone, the second card might depict one medium pink phone, and the third might depict one medium pink fish (see Figure 2). Thus, on all test trials, a test card (e.g., the pink phone) matched one of the cards on one relevant dimension (e.g., shape: the other phone) and matched the remaining card on the other relevant dimension (e.g., color: the other pink item).

All children received the same 12 test trials (presented in the same order), which consisted of two blocks of 6 test trials (see Table 1). Within each block, each of the six possible relevant dimension pairs (e.g., color and shape, shape and size) was presented once in a random order. Each dimension on its own was thus relevant on 6 of the 12 test trials (2 test trials with each of the other possible dimensions), but the same dimension was never presented on more than 2 consecutive trials. The cue of the relevant dimension by which cards matched can be referred to as the *dominant cue* (e.g., pink for color and phone for shape in the example in Figure 2),

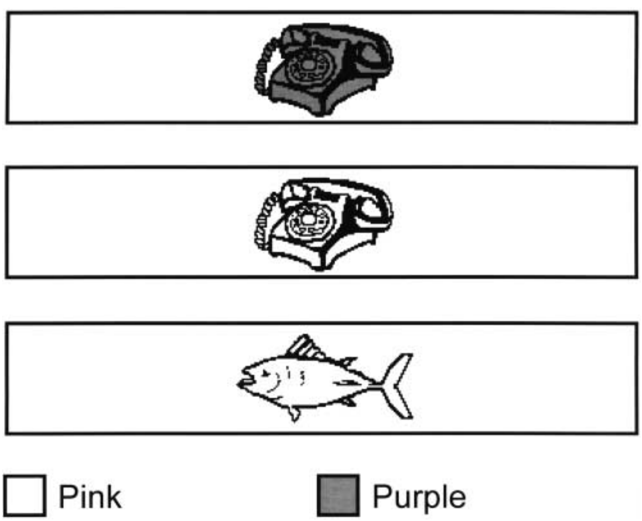

FIGURE 2 Example of cards used in the test trials.

TABLE 1
Descriptive Information About Stimuli Presented on Each Trial as a Function of Block and Trial Number

Block/Trial	*Relevant Dimensions*	*Dominant Cues*	*Nondominant Cues*	*Irrelevant Cues*	*Test Card Placement*
Block 1					
Trial 1	size/number	medium/three	small/one	orange/socks	Card 2
Trial 2	shape/number	socks/one	phone/two	purple/large	Card 1
Trial 3	color/shape	orange/fish	purple/socks	small/one	Card 2
Trial 4	color/number	purple/two	pink/three	fish/medium	Card 3
Trial 5	shape/size	phone/large	fish/medium	pink/two	Card 3
Trial 6	color/size	pink/small	orange/large	phone/three	Card 1
Block 2					
Trial 7	shape/number	fish/three	phone/two	orange/medium	Card 2
Trial 8	shape/size	socks/small	fish/medium	purple/one	Card 1
Trial 9	color/number	orange/one	pink/three	phone/small	Card 1
Trial 10	color/size	purple/medium	orange/large	socks/two	Card 2
Trial 11	size/number	large/two	small/one	pink/fish	Card 3
Trial 12	color/shape	pink/phone	purple/socks	large/three	Card 3

Note. The actual cards presented on each trial can be recreated from the information provided in this table. For example, on Trial 1, children were shown one medium orange pair of socks, three medium orange pair of socks, and three small orange pair of socks, and the test item (i.e., three medium orange pair of socks) was placed in the center position (i.e., Card 2).

whereas the cue of the nonmatching third card can be referred to as the *nondominant cue* of the relevant dimension (e.g., purple for color and fish for shape in the example in Figure 2). Within each block, each cue within each dimension appeared once as the dominant cue of the relevant dimension, once as the nondominant cue of the relevant dimension, and once as the irrelevant cue of an irrelevant dimension (i.e., a dimension that did not vary across the three cards, e.g., medium for size and one for number in the Figure 2 example).

Test card placement was also counterbalanced: The test card appeared as Card 1, Card 2, or Card 3 (from top to bottom). For example, in Figure 2, the placement of the test card is Card 2 because it matches Card 1 on one dimension and Card 3 on the other dimension. Each of the three possible test card placements was used on four trials (two trials within each block; see Table 1), but the same placement was not used on more than two consecutive trials.

Procedure

Children were tested individually in one testing session at their respective day care centers or at the university. They were given the FIST, which took approximately 10 min to complete, in addition to other tasks that were unrelated to this study. Approximately 80% of testing sessions were videotaped. Performance was scored after each trial and confirmed by a second experimenter from videotapes. Furthermore, data from 15 randomly chosen children were rescored by a third experimenter. Scoring reliability was perfect.

Demonstration and criterial trials. Children were introduced to the task by being told that they would see some cards with pictures on them. On the demonstration trial, the experimenter then placed four cards side by side and told the children

> Look! Here's a card, here's another card, here's another card, and here's another card. I'm going to pick two cards that are the same in one way. So I'll pick these two cards [simultaneously pointing to two identical cards, that is, Card 1 and Card 4]. These two cards are the same because they both have one little orange pair of socks on each card. So they're the same. Now I'm going to pick two cards that are the same but in a different way. So I'll pick these two cards [simultaneously pointing to the other pair of identical cards, that is, Card 2 and Card 3]. These two cards are the same because they both have two medium purple fish on each card. That's why they're the same. So these two cards are the same [pointing to the first pair] and these two cards are the same [pointing to the second pair], but see, these two cards here are different from those two cards. You know what? Now it's your turn to show me some cards.

Children were subsequently given two criterial trials. On each criterial trial, they were instructed to "Show me (put your fingers on) two cards that are the same in one way" (Selection 1). Once children responded, they were then asked to "Show me two cards that are the same but in a different way" (Selection 2). Children were not given feedback on their responses because these criterial trials served to ensure that children understood the basic task instructions, including the requirement of selecting two cards, the concept of same, and the concept of same but in a different way. If children erred on even one selection on either of the two criterial trials, they did not receive the task proper.

Test trials. Children who succeeded on both criterial trials were then given 12 test trials. On each of these trials they were also required to make two selections. The instructions were identical to those given in the criterial trials except that test trials involved only three cards instead of four. If children failed to select two cards on either selection within a particular test trial, then the experimenter prompted them on the first occasion in which this occurred (i.e., "Is there another one that's the same as that one?"). Similarly, if children seemed to hesitate after selecting two cards, then the experimenter prompted them on the first instance in which this occurred by asking, "Are you done?"

RESULTS

Performance on Criterial Trials

The percentages of 2-, 3-, 4-, and 5-year-olds who failed at least one of the two criterial trials were 85%, 22%, 12%, and 3%, respectively. A chi-square test confirmed that successful performance on the criterial trials varied as a function of age group, $\chi^2(3, N = 189) = 92.20$, $p < .001$ (assuming an alpha level of .05; see Table 2). Separate Fisher's Exact tests were conducted on each pair of age groups to determine which groups differed from one another. Two-year-olds differed from all other age groups ($ps < .001$), and 3-year-olds only differed from 5-year-olds ($p < .05$). Three-year-olds did not differ from 4-year-olds, and 4-year-olds did not differ from 5-year-olds ($ps > .10$).

Performance on Test Trials

Age-group analyses. Given that only eight 2-year-olds succeeded on the criterial trials—and as a result, completed the task proper—they were excluded from further analyses because they constituted a nonrepresentative sample of 2-year-olds. Consequently, only 3-, 4-, and 5-year-old children who succeeded on the criterial trials were included in the analyses for the test trials. Although the basic design of the experiment was a 3 × 2 (Age × Selection) design with repeated measures on selection, using selection as a repeated measure was problematic because Selection 2 necessarily followed from and, in fact, depended on Selection 1. Consequently, this dependency violated the basic assumption of no carryover effects between treatments in repeated measures designs (Neter, Wasserman, & Kutner, 1990). Therefore, separate analyses were conducted for each selection.

TABLE 2
Number of Children Who Failed or Passed the Criterial Trials as a Function of Age Group

	Performance	
Age Group	*Failed*	*Passed*
2-year-olds	46	8
3-year-olds	11	40
4-year-olds	6	43
5-year-olds	1	34

Inspection of the raw data indicated that the distributions of scores on Selection 1 were negatively skewed, presumably because of ceiling effects on performance. However, as Kirk (1982) noted, "[s]kewed populations have very little effect on either the level of significance or the power of the *F* test for the fixed-effect model" (p. 75). Thus, a one-way analysis of variance (ANOVA) on Selection 1 performance was conducted and a significant main effect of age group was detected, $F(2, 114) = 7.04$, mean square error (MSE) = 2.52, $p < .01$.[1] Effect-size analyses using Cohen's (1988) *f* for multiple-conditions designs revealed that the effect size was large, $f = 0.35$. Post hoc pairwise comparisons using the Tukey honestly significant difference test revealed that 3-year-olds ($M = 9.65$, $SD = 1.53$) did significantly worse than both 4- and 5-year-olds ($M = 10.86$, $SD = 1.51$, and $M = 10.74$, $SD = 1.75$, respectively; *p*s < .05) and that 4- and 5-year-olds did not differ from each other ($p > .10$). Pairwise effect-size comparisons further revealed that the difference between means was large for 3- and 4-year-olds ($d = 0.80$) and moderate for 3- and 5-year-olds ($d = 0.66$). In contrast, as predicted from the ANOVA, the difference between means for 4- and 5-year-olds was negligible ($d = 0.08$).

Given that Selection 2 performance was meaningful as a measure of flexibility only in the context of relatively good performance on Selection 1, data from 3-year-olds were omitted from the analyses of Selection 2 performance because these children did significantly worse on Selection 1 than both 4- and 5-year-olds. A one-way ANOVA on 4- and 5-year-olds' Selection 2 performance revealed a significant difference between age groups, $F(1, 75) = 6.03$, $MSE = 14.57$, $p < .05$. Four-year-olds ($M = 3.79$, $SD = 3.49$) performed significantly worse than 5-year-olds ($M = 5.94$, $SD = 4.20$) on Selection 2 despite equivalent performance on Selection 1. This difference was moderate in terms of effect size, $d = 0.56$. Figure 3 displays the mean number of correct trials (and standard errors) that children obtained on each selection as a function of age group.

Response pattern analyses. Selection 1 responses were classified into five mutually exclusive (and exhaustive) categories. Responses were classified as *correct pair* if children selected a correct pair of cards (note that two correct pairs were possible for Selection 1). Incorrect responses were classified as *wrong pair* if children selected the (only) wrong pair of cards (e.g., the purple phone and the pink fish in the example in Figure 2). The other three response categories occurred when children selected either more or fewer than two cards, including selecting (a) *all cards,* (b) *one card,* or (c) *no cards*.

Table 3 illustrates the percentages of Selection 1 (and Selection 2) responses that were of each type summed across all trials and across all children within each

[1]Analyses reported in this article were first conducted with sex as one of the predictor variables. However, no main effect or interactions involving sex were detected. For simplicity, sex was dropped from all analyses and the results are presented without it.

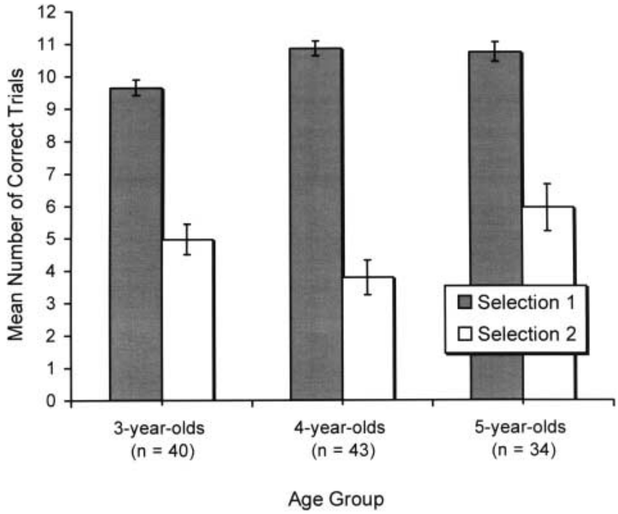

FIGURE 3 Mean numbers (and standard errors) of correct trials as a function of age group and selection ($N = 117$). (Note that the mean and standard error of 3-year-olds' Selection 2 responses are also presented despite the fact that 3-year-olds were not included in the analysis on Selection 2 performance because of their relatively poor Selection 1 performance.)

age group. Although no inferential statistics could be conducted on these data because observations were not independent of each other, it is worth noting that 3-year-olds selected the wrong pair on Selection 1 almost three times more often than older children. This finding is consistent with the results of the ANOVA, which revealed that 3-year-olds had more difficulty on Selection 1 than the older children. Also of interest is the finding that on Selection 1, the percentages of other errors (i.e., all cards, one card, and no cards) were relatively low (less than 5% combined for each age group) and were similar across all age groups.

Selection 2 responses were also classified into separate categories. In addition to the five categories used for Selection 1 responses, two other response categories were added. On Selection 2, it was also possible for children to err by selecting the *same pair* of cards that they selected on Selection 1. Furthermore, one-card responses were classified into two separate categories: These included *remaining-card* responses, which occurred when children selected only the card that they had not selected on Selection 1, and *other* one-card responses, which included all other instances of one-card responses. Hence, response categories for Selection 2 included the following mutually exclusive and exhaustive categories: (a) correct pair, (b) wrong pair, (c) same pair, (d) all cards, (e) remaining card, (f) other one card, and (g) no cards.

The Selection 2 data were different from Selection 1 data in several respects. As shown in Table 3, unlike Selection 1 in which children tended to err by selecting

TABLE 3
Overall Percentages for Each Possible Response Category Across Children as a Function of Age Group and Selection

	Responses Categories						
Age Group	*Correct Pair*	*Wrong Pair*	*Same Pair*	*All Cards*	*One Card*	*Remaining Card*	*No Cards*
3-year-olds[a]							
Selection 1	80.4	16.9	—	0.2	2.5	—	0.0
Selection 2	41.5	20.8	12.9	0.2	1.7	22.7	0.2
4-year-olds[b]							
Selection 1	90.5	6.0	—	0.2	3.1	—	0.2
Selection 2	31.6	17.4	6.8	0.2	1.9	40.5	1.6
5-year-olds[c]							
Selection 1	89.5	5.9	—	0.2	2.7	—	1.7
Selection 2	49.5	15.4	10.5	0.0	2.9	16.9	4.7

Note. The same pair and remaining card response categories applied only to Selection 2 responses, and one card responses for Selection 2 consisted of other one-card responses (for definitions on each response category, see the Response pattern analyses subsection of the Results section). Each child within each age group contributed 12 responses for each selection.

[a]$n = 40$. [b]$n = 43$. [c]$n = 34$.

the wrong pair, the most common types of error on Selection 2 were remaining-card responses—particularly for the 4-year-olds who erred in this manner approximately twice as often as either the 3- or 5-year-olds. The next most frequent types of incorrect responses for all age groups were wrong pair and same pair, whereas the other types of incorrect responses occurred infrequently. Morever, unlike Selection 1, it was the 4-year-olds, not the 3-year-olds, who erred on a greater proportion of trials than the other age groups.

DISCUSSION

Several aspects of executive function have been well studied in preschoolers (e.g., rule use, inhibitory response control, working memory, planning, and error evaluation; see Dempster, 1992, 1993; Harnishfeger & Bjorklund, 1993; Zelazo et al., 1997; and Zelazo & Jacques, 1996, for reviews), but tests assessing abstraction and cognitive flexibility have been lacking. In our study, 2-, 3-, 4-, and 5-year-olds were tested on a new task, the FIST, which provides a relatively straightforward measure of these aspects of executive function in preschoolers. Important and meaningful age-related differences emerged in this study across the entire age range tested.

On one hand, the majority of 2-year-olds showed no evidence of understanding basic task instructions, as revealed by their poor performance on criterial trials. On the criterial trials, children were given instructions that were identical to those given on the FIST; however, children did not have to detect a specific dimension by which to match a pair of cards (i.e., no abstraction component) because matching cards were identical on all dimensions. Furthermore, they did not have to select a particular card in multiple ways (i.e., no flexibility component) because there were four cards instead of three, and therefore, each card needed to be selected only once. Hence, 2-year-olds' difficulty on the criterial trials cannot be attributed to limitations in abstraction or cognitive flexibility per se.

On the other hand, findings from the FIST itself converge to provide a clear picture of changes in abstraction and flexible thinking over the 3- to 5-year period. Most 3-year-olds succeeded on criterial trials but performed worse than the older age groups on Selection 1. To perform well on the test trials, children needed to detect or abstract dimensional information because matching cards were nonidentical. In contrast, the matching cards in the criterial trials were identical and therefore, it was not necessary for children to abstract dimensional information or match cards on the basis of a particular dimension. More specifically, in the criterial trials it was possible for children to rely on a simpler perceptual-matching strategy in that they could match cards according to overall similarity without necessarily identifying a specific dimension. Ample evidence exists in the literature to support the notion that young children do indeed use this kind of simpler strategy (see Gentner & Rattermann, 1991; Smith, 1989; Zelazo & Jacques, 1996, for reviews). The relatively poor performance of 3-year-olds on Selection 1 compared to their own good performance on the criterial trials (and to the performance of the older age groups on Selection 1) suggests that they had difficulty correctly detecting a dimension that was common to two cards when these cards were not identical to each other. Thus, 3-year-olds' difficulty on this task appears to be due primarily to difficulty abstracting out a relevant dimension.

In contrast, 4-year-olds performed as well as 5-year-olds on Selection 1 of the test trials, suggesting that they correctly detected how two nonidentical cards were identical along one dimension. In other words, they appeared to do well on the abstraction component of the task. However, 4-year-olds' relatively poor performance on Selection 2 of these test trials compared to 5-year-olds is consistent with the suggestion that it was the requirement that they flexibly select one of the cards (e.g., the pink phone in Figure 2) according to two dimensions that rendered the task difficult for them. Indeed, when they erred on Selection 2, they tended to do so by selecting the remaining card alone (see following for further discussion of remaining-card responses). In contrast, on Selection 1 most errors occurred because children selected the wrong pair, demonstrating that they at least understood to some degree that they were expected to select two cards. Hence, despite understanding that they needed to select two cards—as evidenced by their good performance on the criterial trials and on Selection 1 of the test trials—4-year-olds tended to select only one card on Selection 2. A substantial minority also erred by selecting the same pair twice. Selecting the remaining card alone raises the possibility that children failed to see how the test card (i.e., the card that needed to be selected twice, e.g., the pink phone in Figure 2) matched the remaining card on a different dimension. They may have perseverated in thinking of the test card in only one way (i.e., according to the first dimension by which they selected it) and consequently refused to select it with the only card that was left after they had selected cards for Selection 1.

Note, however, that the failure to select the test card twice could also be due to a misunderstanding of the task requirements. The criterial trials required children to understand the basic task instructions but did not require them to abstract or represent flexibly dimensional information. Although inclusion of these trials was necessary to determine at which age children understood task instructions, the design of the criterial trials may have inadvertently led children to believe that each card ought to have been selected only once, given that correct responses in the criterial trials never required selecting any given card more than once. Of course, a combination of both of these explanations of remaining-card responses is also possible. For example, in the test trials children may have failed to see how the test card matched the remaining card on a second dimension (i.e., cognitive inflexibility), and because of the lack of a fourth card, they may have opted to select the remaining card alone. On this account, their difficulty stemmed from cognitive inflexibility, but how their difficulty manifested itself depended on task variables (e.g., the inclusion of the criterial trials). The results of a new experiment with the FIST suggest that this may in fact be the case (Jacques, Zelazo, Lourenco, & Sutherland, 2001, Experiment 1).

Criterial trials were not given in Jacques et al. (2001, Experiment 1). Instead, children were given a demonstration trial and two practice trials with feedback that were identical in format to the stimuli presented in the test trials. Thus, children were explicitly shown in these preliminary trials how to select two items on both selections (and therefore to select at least one of the items twice). Findings in Jacques et al. (2001) from a no-label condition (i.e., a condition that most closely resembles the procedure used for administering the FIST in this experiment) replicated those of this experiment. That is, on Selection 1, 3-year-olds did worse than 4- or 5-year-olds (who did not differ), whereas on Selection 2, 4-year-olds did worse than 5-year-olds. However, on Selection 2, children in that experiment tended to err by selecting either the wrong pair or the same pair twice rather than by selecting the remaining item alone as in this experiment. The results of the current study together with those of Jacques et al. (2001, Experiment 1) suggest that certain task variables (e.g., inclusion of criterial trials) only affect how children's underlying difficulty with switching flexibly between dimensions manifests itself and not whether they experience difficulty.

Our interpretation of the findings of the current study also rests on the assumption that Selection 1 responses measure abstraction, whereas Selection 2 responses measure cognitive flexibility. Of course, as mentioned previously, abstraction is also necessary for selecting matching pairs on Selection 2, although one might reasonably expect that if children were to have difficulty with the abstraction component on Selection 2 then they would also have difficulty on Selection 1. However, if children were to always select the dimension that they found easier to abstract on Selection 1, then poor performance on Selection 2 might arise from difficulty abstracting a more difficult—or less preferred—dimension, instead of arising from difficulty with switching between dimensions per se (i.e., the cognitive flexibility component).

Although the results of our experiment do not distinguish between these two possibilities, the results of a more recent experiment with the FIST (Jacques et al., 2001, Experiment 2) indicate that even when the relative abstraction difficulty of each selection is controlled, 4-year-olds continue to exhibit difficulty on Selection 2, supporting the notion that Selection 2 responses do indeed provide an index of cognitive flexibility. More specifically, in the Jacques et al. (2001) experiment, it was the experimenter who always selected items on Selection 1 in a predetermined and counterbalanced manner, and 4-year-olds were assessed only on their Selection 2 performance. As a result, the experimenter should have selected items on the basis of children's preferred dimension on about half of the trials and on the basis of their less preferred dimension on the other half. If Selection 2 responses measure their difficulty with abstracting less preferred dimensions, then children ought to have done well on Selection 2 on those trials in which the experimenter selected items on the basis of their less preferred dimension (because selecting items according to their preferred dimension was still possible). Thus, their overall performance on Selection 2 should have been better than the performance of children asked to select items themselves on Selection 1. In contrast, if Selection 2 responses measure their difficulty with switching between dimensions, then children's performance ought to have been similar to that of children who were asked to select items themselves on Selection 1. As it happened, the performance of the 4-year-olds in Jacques et al. (2001) was much like that of the 4-year-olds in our experiment, suggesting that Selection 2 responses do in fact provide a means of assessing children's ability to switch between dimensions rather than their ability to detect less preferred dimensions.

The results from this task converge well with those from another task that superficially resembles the FIST. In the Dimensional Change Card Sort (e.g., Frye et al., 1995; Zelazo et al., 1996; see Zelazo & Frye, 1997, and Zelazo & Jacques, 1996, for reviews) children are presented with two target cards (e.g., a red car and a blue flower) that remain visible throughout the task, and children are asked to sort test cards (e.g., blue cars and red flowers) that match one of the target cards on one dimension (e.g., shape) and match the other target card on the other dimension (e.g., color). Children are first told explicitly to sort test cards by one dimension for a number of trials and then they are asked to switch and sort cards by the other dimension. Despite being told the relevant rules on each trial (e.g., "Red ones go here and blue ones go there"), the majority of 3-year-olds sort perseveratively in the postswitch phase. In contrast, the majority of 4- and 5-year-olds switch and sort test cards according to the postswitch rules. Furthermore, findings from several studies (e.g., Jacques et al., 1999; Zelazo et al., 1996) suggest that poor performance on this task results from limitations in flexible thinking and not from difficulty with inhibitory response control.

In addition to differences in the procedures for administering the tasks, one important difference between the FIST and the Dimensional Change Card Sort is that the FIST is an inductive as opposed to a deductive task. That is, on the Dimensional Change Card Sort, children are told the dimension by which to sort, whereas on the FIST, children are required to identify the relevant dimension themselves. Thus, although both tasks require that children use dimensional information and that they be able to switch flexibly between different dimensions, only the FIST requires that children abstract relevant dimensional information for themselves. Perhaps the *décalage* between 4-year-olds' poor performance on the FIST and their relatively good performance on the Dimensional Change Card Sort reflects the added requirement of the FIST that children both abstract the relevant dimension (as opposed to being told which dimension to use) and use that information flexibly. Findings from this study reveal that the abstraction component of the FIST is not problematic for 4-year-olds, as evidenced by their good performance on Selection 1. Similarly, findings using the Dimensional Change Card Sort indicate that cognitive flexibility requirements alone are not problematic for 4-year-olds. Instead, it appears it is the dual requirement that children both abstract and switch flexibly between dimensions that renders the FIST particularly difficult for 4-year-olds. Future work should match more closely the procedures for administering both of these tasks to test this hypothesis.

In sum, the results revealed that 3-year-olds experienced difficulty in abstracting the relevant dimension on Selection 1, lending further support to the notion that preschoolers respond to relatively concrete aspects of stimuli (e.g., Inhelder & Piaget, 1959/1964; Kendler, 1972; Smith, 1989; Vygotsky, 1978, 1934/1986; but see DeLoache, Miller, & Pierroutsakos, 1998, for a review of work suggesting that young children are capable of abstracting relevant information when external feedback is provided). The majority of 4-year-olds, on the other hand, did well on Selection 1. However, the findings indicate that they had difficulty switching between dimensions on this inductive task, providing further evidence for the development of cognitive flexibility during the preschool years (e.g., Gerstadt et al., 1994; Inhelder & Piaget, 1959/1964; Jacques et al., 1999, 2001; Zelazo et al., 1996). Finally, 5-year-olds performed significantly better than 4-year-olds on Selection 2, although their performance was not at ceiling. The findings of this study also converge well with findings obtained using simpler deductive tasks that assess flexible thinking (e.g., Diamond & Taylor, 1996; see Zelazo & Jacques, 1996, for a review). Future studies designed to determine how the FIST relates to other tasks assessing different aspects of executive function and how it might relate to more general abilities such as language development (see Jacques et al., 2001) will undoubtedly inform our understanding of the processes necessary for successful performance on the FIST. Longitudinal designs would be particularly useful for understanding the exact nature of the relations between performance on the FIST and more general abilities. Furthermore, although the FIST has yet to be used with groups of children with different brain dysfunctions, it promises to provide important insights into the development of specific aspects of executive function that are often conflated in traditional measures of problem solving.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the National Sciences & Engineering Research Council of Canada (NSERC) to Philip David Zelazo, and an NSERC Postgraduate Fellowship and Ontario Graduate Scholarship to Sophie Jacques. We thank Victoria Orekhovsky, and especially Jodie M. Burton, for invaluable assistance with participant recruitment and data collection. We are also grateful to Stella Felix Lourenco, Stuart Marcovitch, and Ulrich Müller for providing helpful comments on a previous draft. This work was submitted as part of a doctoral dissertation to the University of Toronto. A portion of this work was also presented at the 105th Annual Convention of the American Psychological Association, Chicago, IL, August 1997.

REFERENCES

Alp, I. E. (1994). Measuring the size of working memory in very young children: The Imitation Sorting Task. *International Journal of Behavioral Development, 17,* 125–141.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology, 39,* 15–22.

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology, 20,* 607–618.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386–404.

Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology, 8,* 219–228.

Chelune, G. J., & Thompson, L. L. (1987). Evaluation of the general sensitivity of the Wisconsin Card Sorting Test among younger and older children. *Developmental Neuropsychology, 3,* 81–89.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Delis, D. C., Squire, L. R., Bihrle, A., & Massman, P. (1992). Componential analysis of problem-solving ability: Performance of patients with frontal lobe damage and amnesic patients on a new sorting test. *Neuropsychologia, 30,* 683–697.

DeLoache, J. S., Miller, K. F., & Pierroutsakos, S. L. (1998). Reasoning and problem solving. In W. Damon (Series Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 801–850). New York: Wiley.

DeLoache, J. S., Sugarman, S., & Brown, A. L. (1985). The development of error correction strategies in young children's manipulative play. *Child Development, 56,* 928–939.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12,* 45–75.

Dempster, F. N. (1993). Resistance to interference: Developmental changes in a basic processing mechanism. In M. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development* (Vol. 1, pp. 3–27). New York: Springer-Verlag.

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.

Drasgow, J., & Feldman, M. (1957). Conceptual processes in schizophrenia revealed by the Visual–Verbal Test. *Perceptual and Motor Skills, 7,* 251–264.

Feldman, M. J., & Drasgow, J. (1951). A Visual-Verbal Test for schizophrenia. *Psychiatric Quarterly, 25*(Suppl.), 55–64.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483–527.

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13,* 343–359.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 225–277). Cambridge, England: Cambridge University Press.

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.

Gordon, A. C. L., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68,* 70–83.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38,* 404–411.

Harnishfeger, K. K., & Bjorklund, D. F. (1993). The ontogeny of inhibition mechanisms: A renewed approach to cognitive development. In M. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development* (Vol. 1, pp. 28–49). New York: Springer-Verlag.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation* (E. A. Lunzer & D. Pepert, Trans.). New York: Harper & Row. (Original work published 1959)

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental Psychology, 35,* 770–780.

Jacques, S., Zelazo, P. D., Lourenco, S. F., & Sutherland, A. (2001). *Age-related changes in preschoolers' performance on the Flexible Item Selection Task (FIST): The roles of labeling and abstraction in the development of cognitive flexibility.* Manuscript in preparation.

Kendler, T. S. (1972). An ontogeny of mediational deficiency. *Child Development, 43,* 1–17.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem solving and planning processes in preschool children. *Cognitive Psychology, 13,* 113–148.

Levine, B., Stuss, D. T., & Milberg, W. P. (1995). Concept generation: Validation of a test of executive functioning in a normal aging population. *Journal of Clinical and Experimental Neuropsychology, 17,* 740–758.

Luria, A. R. (1959). The directive function of speech in development and dissolution. Part I. Development of the directive function of speech in early childhood. *Word, 15,* 341–352.

Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behaviour* (J. Tizard, Ed.). New York: Pergamon.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9,* 100–111.

Milner, B., & Petrides, M. (1984). Behavioural effects of frontal-lobe lesions in man. *Trends in Neurosciences, 7,* 403–407.

Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental designs* (3rd ed.). Homewood, IL: Irwin.

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

Siegel, S. M. (1957). Discrimination among mental defective, normal, schizophrenic and brain damaged subjects on the Visual-Verbal concept formation test. *American Journal of Mental Deficiency, 62,* 338–343.

Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 146–178). Cambridge, England: Cambridge University Press.

Stuss, D. T., Benson, D. F., Kaplan, E. F., Weir, W. S., Naeser, M. A., Lieberman, I., & Ferrill, D. (1983). The involvement of orbitofrontal cerebrum in cognitive tasks. *Neuropsychologia, 21,* 235–248.

Stuss, D. T., Eskes, G. A., & Foster, J. K. (1994). Experimental neuropsychological studies of frontal lobe functions. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 149–185). Amsterdam: Elsevier.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press. (Original work published 1934)

Welsh, M. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi Disk-Transfer Task. *Cognitive Development, 6,* 59–76.

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.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1,* 198–226.

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. Stamenov (Ed.), *Language structure, discourse, and the access to consciousness* (pp. 113–153). Amsterdam: Benjamins.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11,* 37–63.

Zelazo, P. D., & Jacques, S. (1996). Children's rule use: Representation, reflection and cognitive control. In R. Vasta (Ed.), *Annals of child development* (Vol. 12, pp. 119–176). London: Jessica Kingsley.