WORKING MEMORY CAPACITY AND AGE AS DETERMINANTS OF PERFORMANCE IN THE BALLS AND BOXES PUZZLE

SU XIAO

Working Memory Capacity and Age

as Determinants of Performance in the Balls and Boxes Puzzle

by

Su Xiao

A thesis submitted to the School of Graduate Studies in partial fulfillment of the

Master of Science

Department of Psychology

Memorial University of Newfoundland

November, 2007 Submitted

St. John's Newfoundland

Abstract

Reber and Kotovsky (1997) claimed that even though learning to solve the Balls and Boxes puzzle is implicit, it is slowed by a secondary task, thus suggesting that implicit learning requires attentional capacity. In the present study, this suggestion and the degree to which implicit learning can be attributed to age-correlated changes were tested by comparing individuals differing in test-defined working memory (WM) capacity. Retrospective verbal reports, a move-selection test, and Trial 2 performance data all indicated that participants were unaware of their knowledge of the puzzle, suggesting implicit learning. However, speak-span scores did not correlate with performance measures on either the learning or transfer trial. It appears that in the absence of a secondary task, WM capacity did not affect learning or transfer in the Balls and Boxes puzzle. Moreover, inconsistent with Reber's (1992, 1993) age-independent assumption, substantial developmental changes on performance were found when the children in the present study were compared to the adults in Reber and Kotovsky's (1997) study.

Keywords: Implicit learning, working memory, age, children, Balls and Boxes puzzle

Acknowledgements

Preparation of this thesis was supported by Grant OGP0002017 (to F. Michael Rabinowitz) from the Natural Science and Engineering Research Council of Canada. I am grateful to the Eastern School District of St. John's, the principals, parents, and children of Macpherson, St. Andrew's, Rennie's River, and Bishop Field Schools for participating; to Kun Li for writing the program of the Balls and Boxes puzzle; to Stephen Chung for conducting part of the experiment; and to Michael Rabinowitz, Malcolm Grant and Catherine Penney for their helpful suggestions.

List of Figures

List of Tables

Table of Contents

Introduction

In everyday life, we often solve problems and make decisions following rules that we can state. When using a recipe, we can describe what we are doing and how we are doing it. Learning of this kind, that can be articulated or demonstrated on demand, is referred to as explicit learning (Berry & Dienes, 1993; Cleeremans, 1993; Kirkhart, 2001). However, a good deal of our knowledge and skills are not describable. For example, most of us understand and produce grammatical utterances in our native language without being able to articulate the rules we are following. Non-articulated learning of this kind, that can take place in an incidental manner, is referred to as implicit learning (Reber, 1989). It can be characterized by behavioral sensitivity to the structure of the environment and lack of awareness of this sensitivity (e.g., Berry, 1997; Berry & Dienes, 1993; Cleeremans, 1993, 1997; Frensch & Rünger, 2003; Kirkhart, 2001; Reber, 1993). Although it seems clear that implicit learning needs to be contrasted with learning that is not implicit, defining and operationalizing implicit learning remains a challenge (Frensch, 1998).

In this paper, the effects of working memory capacity and age on implicit learning are assessed. It is usually assumed that implicit learning is resource free (Nissen & Bullemer, 1987), therefore, the rate of implicit learning should not depend on either working memory capacity or age. In the remainder of this paper, the relationships among working memory, age, and implicit learning are reviewed. A rationale for employing the Balls and Boxes puzzle, a problem-solving task used to study implicit learning in the present experiment, is then provided. The experimental details follow.

Implicit learning has been studied in a variety of tasks including covariation learning (Musen & Squire, 1993), artificial grammar (Reber, 1989), sequential reaction time (SRT) (Nissen & Bullemer, 1987), and problem-solving (Balls and Boxes Puzzle, Reber & Kotovsky, 1997; Tower of Hanoi, Zanga, Richard, & Tijus, 2004). Whether learning in these tasks operates without conscious knowledge has been questioned (Perruchet, Chambaron, & Ferrel-Chapus, 2003; Shanks & St. John, 1994). Shanks and St. John (1994) introduced two criteria to ascertain whether implicit learning had been established: (a) the information criterion, the information assessed by awareness tests must be similar to the information responsible for performance changes; and (b) the sensitivity criterion, the awareness tests must be sensitive to relevant conscious knowledge.

Attention, Working Memory and Implicit Learning

Working memory is defined as "a system of processes and stores used to maintain information during processing" (p.341, Hambrick & Engle, 2002) or "a resource-limited processing construct for executing the computations necessary to execute problem-solving behavior" (p.195, Reber & Kotovsky, 1997). The relationship between attention and working memory plays a central role in many cognitive theories. There is a consensus that working memory is of limited capacity

(Oberauer, 2005), attention is a prerequisite to memory acquisition (Nissen & Bullemer, 1987), and the ability to control attention is an important function of working memory (Barrett, Tugade, & Engle, 2004; Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003; Conway, Kane, Buting, Hambrick, Wilhelm, & Engle, 2005).

Working memory has been investigated in two ways: by comparing individuals who differ in test-defined working memory capacity or by experimentally varying working memory load in a primary or in a secondary task. Evidence clarifying the relationship between working memory capacity, working memory load, and cognitive performance has been accumulated following the seminal work of Baddeley and Hitch (1974, cited in Conway & Engle, 1996). In general, in difficult tasks, if working memory is resource-limited, explicit learning rate should correlate positively with memory capacity and negatively with memory load. Since implicit learning, unlike explicit learning, is assumed to proceed without making any demands on attentional resources (Berry & Dienes, 1993; Reber, 1993), this type of learning should be unrelated to both working memory capacity and memory load.

Several investigators have manipulated memory load in implicit learning tasks using college students as participants (e.g., Cohen, Ivry & Keele, 1990; Nissen & Bullemer, 1987; Reber & Kotovsky, 1997). A secondary task, typically tone-counting, was used to increase memory load. Surprisingly, negative correlations between increased memory load and learning rate were found in most reports (e.g., Nissen &

Bullemer, 1987; Reber & Kotovsky, 1997; Shanks & Channon, 2002). Since the relationship between working memory capacity and implicit learning not been investigated, this was the focus in the current study.

The sequential reaction time (SRT) task is used to investigate implicit learning. In this task (Nissen & Bullemer, 1987), an asterisk appears at one of four horizontally arranged locations on a display. Participants are instructed to press the key located below the activated asterisk as quickly as possible. Following a correct button press, the asterisk is extinguished and a new asterisk appears at another location. The sequence in which the asterisks appear is varied across participants. In the repeating condition, the sequence is fixed. In the random condition, the location of the stimulus on each trial is quasi-randomly determined with the constraint that a position can not repeat on successive trials. Participants usually are not informed of sequence type. Three criteria are indicators of implicit learning: (a) reaction times decline reliably faster with the repeating sequence than with the random sequence during training; (b) reaction times increase significantly when the repeating sequence is replaced by a random sequence; and (c) participants appear to be unaware of the repeating sequence following training (Nissen & Bullemer, 1987).

Nissen and Bullemer (1987) presented four continuous blocks, each containing 10 repetitions of a 10-trial sequence (i.e., 4231324321, designating the four locations as 1, 2, 3, and 4 from left to right) in the repeating condition. The end of one 10-trial sequence and beginning of the next was not marked in any way. Nissen and Bullemer (1987) tested three groups of participants: the single-repeating group, the dual-repeating group, and the dual-random group. The single-repeating group did not experience a secondary task while the other two groups performed the SRT task and a tone-counting task simultaneously. A high-pitched or low-pitched tone was emitted during the intervals between asterisk offset and onset. Participants were required to maintain an accurate running count of the number of low tones and to report the count at the end of each block. Nissen and Bullemer (1987) found slower responses in the dual-task conditions than in the single-repeating condition. Furthermore, comparable decreases in the response times of the dual-repeating and dual-random groups were obtained across Blocks 1-4. Thus, the improvement demonstrated by the dual-repeating group reflected dual-task practice rather than sequence learning.

After four training blocks, sequential knowledge was assessed using a single-generation task (without a secondary task) for two blocks of 100 trials. On each trial, an asterisk appeared at one of four locations in the sequence used in the repeating conditions. Participants were instructed to press the key corresponding to the location in which they thought the next asterisk would appear and were informed that accuracy rather than speed was of interest. Transfer from the training to the generation task was apparent in the data of the single-task, but not the dual-task, training groups. Therefore, learning did not occur if a secondary task was added during training.

In a second experiment, Nissen and Bullemer (1987) further investigated whether participants learn a repeating sequence under dual-task conditions by assessing transfer using a single task. Two groups experienced four training blocks of a repeating sequence under the dual-repeating condition. On the single-task transfer blocks, one of the groups was presented the same repeating sequence, and the other group was presented a quasi-random sequence. The mean latencies and the mean accuracy across the 10-trial sets in Block 5 for these two groups were comparable. Nissen and Bullemer (1987) concluded that the tone-counting task reduced the amount of WM capacity that was available for performing the SRT task. Since sequence learning was obtained in the single-repeating condition, but not the dual-repeating condition, it appears that learning is dependent on WM capacity.

Cohen, Ivry and Keele (1990) assessed dual-task SRT performance using unique, ambiguous, and hybrid sequences. The unique sequences involved five signal positions, none of which was repeated, with the stimulus location on one trial predicting the stimulus location on the next trial. Four different versions were used (i.e., 12354, 13425, 14532, or 15243) in blocks of 100 trials with 20 cycles per block. An ambiguous sequence was one in which the location on trial n did not uniquely determine the location on trial n+1. Using three locations, the constraint of no unique association mandated that the sequence must contain at least six elements. Three ambiguous sequences were used: 123132, 123213, or 132312. The hybrid sequence (e.g., 142312) had two unique associations (position 4 is followed only by position 2, and position 3 is followed only by position 1) and two ambiguous associations (23 and 21; 14 and 12). Six versions were used: 123243, 123134, 143132, 142312, 132412, and 423213. Each ambiguous and hybrid sequence cycled 20 times in 120-trial blocks. Fourteen blocks of 100- or 120-trials were presented to participants consisted of two practice blocks with random sequences, eight training blocks with one of the three fixed sequences, two more random blocks, and two final fixed blocks with the training sequence. A secondary tone-counting task was presented concurrently in both the practice and training phases. All participants were required to count the number of high-pitched tones, ignoring the low-pitched ones, and report the number as accurately as possible at the end of each block. Following the 14 blocks, participants performed a single-generation task (without a secondary task).

Cohen et al. (1990) found that both unique and hybrid sequences could be learned in the presence of the tone-counting task. Reaction times decreased with practice, increased about 90 ms when a random sequence was presented, and rebounded on the shift back to the repeating sequence. By contrast, the tone-counting task interfered with learning of the ambiguous sequences. The effects of changing from repeating sequence to random and back was not significant (about 20 ms). None of the groups appeared to be aware of the repeating sequence on the generation task. In a second experiment, Cohen et al. (1990) assessed the ability of the ambiguous group to learn the sequence under single-task condition. Reaction times increased about 80 ms with the switch from the repeating sequence to a random sequence; the shift back to the repeating sequence produced an improvement of 100 ms. The difference between reaction times on random blocks and the average reaction time on the pre-shift and post-shift blocks was significant (p <.01). Thus, ambiguous sequences were learned only in the single-task condition (Cohen et al., 1990).

Cohen et al. (1990) confounded sequence length, block length, and number of responses with sequence types. First, sequences lengths were different across the three sequence types: five for the unique sequence and six for the others. Neither of the sequence lengths was comparable to the 10-trial sequence used by Nissen and Bullemer (1987). Second, all sequences cycled 20 times; however, due to the different sequence lengths, the block lengths also differed with 100 trials per block for unique sequences and 120 trials for the other two sequences. Third, the number of response alternatives was not equal. The unique group experienced a five-position task; four positions were used with the hybrid group, and three with the ambiguous group. Any of these confounds might have produced Cohen et al.'s results.

Nissen and Bullemer (1987) argued that tone-counting affected implicit sequence learning by reducing the amount of attention available while Cohen et al. (1990) found the interference effect only with ambiguous sequences. In contrast, Frensch, Lin, and Buchner (1998) argued that the secondary task affects concurrent performance rather than learning. They conducted four experiments to test their hypothesis. In Experiments 1a and 1b, six different repeating 9-trial hybrid sequences (e.g., 135641625) were presented. Participants experienced different combinations of single-task (ST) and dual-task (DT) training across seven blocks. The repeating sequence was replaced by random sequences on Blocks 8 and 9, followed by the original repeating sequence on Blocks 10 and 11. When participants had completed Block 11, a recognition test containing 72 sequences of varying lengths (i.e., 3 to 7 items) was presented. Half of these sequences had appeared in the training phase. Participants were asked to judge whether they had seen the patterns during training.

In Experiment 1a, participants were trained first under DT and second under ST conditions. In the 2-DT/5-ST (4-DT/3-ST, 6-DT/1-ST) condition, 2 (4, 6) dual-task blocks of trials were followed by 5 (3, 1) single-task blocks of trials. In Blocks 8 and 9, random sequences were presented under ST conditions. After eliminating the explicit learners, whose recognition scores were greater than or equal to 90%, all the implicit learners improved their performance across the training blocks. However, it was not clear whether participants obtained knowledge about sequential associations, learned to process two tasks simultaneously, or both.

In Experiment 1b, DT and ST trial blocks were exchanged: 2-ST/5-DT, 4-ST/3-DT, and 6-ST/1-DT. In Blocks 8 and 9, random sequences were presented under DT conditions. After eliminating the explicit learners, Frensch, et al. (1998)

showed that performance reached asymptote on the first ST block. When a secondary task was added, performance deteriorated at first, then improved across blocks, but never to the level as in any the ST block. Therefore, during DT training, participants mainly learned how to process the two tasks simultaneously.

In Experiment 2a, French et al. (1998) used Cohen et al.'s (1990) 6-trial, 4-position ambiguous sequences; whereas in Experiment 2b, they presented participants with 6-trial, 6-position unique sequences. French et al. (1998) confounded training and transfer conditions such that participants who experienced ST training were tested under ST conditions first and DT conditions second, whereas those who experienced DT training were tested under DT conditions first and ST conditions second. It appeared that the groups trained under ST conditions learned the ambiguous sequences although they were more difficult to learn than were the unique sequences. However, learning was masked by the test conditions because participants obtained higher learning scores when tested under ST conditions.

Four features of Frensch et al.'s methodology in Experiments 1a and 1b might have reduced or masked the effects of a secondary task on implicit SRT learning. First, Frensch et al. gave all of their groups both single- and dual-task training in a within-subject design rather than giving one group only single-task training and another group only dual-task training in a between-subject design. Experiencing dual-task training might have interfered with subsequent single-task learning. Second, the training conditions of the most extreme groups (i.e., 2-DT/5-ST vs. 6-DT/1-ST) differed in only four blocks of trials. Third, rather than using the common criterion of 10% (e.g., Cohen et al., 1990), Frensch et al. (1998) included the data of all participants whose average tone-counting errors on the dual-task training blocks were less than 20%. Thus, some of the included participants might have attended minimally to the secondary task. Fourth, some of the training sequences (e.g., 135641625) contained no reversals (e.g., 121) whereas the random sequence presented in the eighth and ninth blocks did. If participants learned that training sequences contained no reversals, and thus, knew that the target would not appear in the two preceding locations, then the reaction times would be expected to be particularly slow on trials containing reversals in the quasi-random blocks. Hence, the transfer scores Frensch et al. obtained may have been inflated across all conditions.

Changing some of the methodological features, Shanks and Channon (2002) conducted two experiments to investigate Frensch et al.'s (1998) hypothesis that dual-task testing conditions adversely affect performance rather than learning. All participants were presented fourteen 96-trial blocks followed by a 96-trial generation task. On Blocks 1 to10, both Single-Repeating and Dual-Repeating groups were trained with a fixed 12-trial sequence while the Dual-Nonrepeating group experienced quasi-random trials. All groups experienced the fixed sequence under ST conditions on Blocks 11, 13, and 14. On Block 12, rather than switching participants to a quasi-random sequence, they were switched to a different fixed sequence that was structurally related to the one used in the training phase under ST conditions.

Unlike French et al. (1998, Experiments 1a and 1b), Shanks and Channon (2002) found a substantial effect of training conditions. After participants who made more than 10% tone-counting errors were eliminated, the index of sequence knowledge, operationalized as the difference between the reaction time on Block 12 and the average reaction time across Blocks 11 and 13, were similar in the dual-task groups. However, their scores were smaller than those of the Single-Repeating group, and this difference appeared to be unrelated to explicit knowledge.

Shanks, Rowland, and Ranger (2005) explored the extent to which sequence knowledge was implicit. They introduced probabilistic sequences to reduce the likelihood of explicit learning. Learning of probabilistic sequences is indexed by comparing reaction times on high and low probability target locations (Cleeremans & Jimenez, 1998). Shanks et al. (2005) used two structurally identical 12-trial sequences containing reversals: A=242134123143 and B=343124132142. For both sequences, any item could follow another but bigrams uniquely determined the subsequent item. During the fourteen 100-trial training blocks, which began at a random point in the sequence for each block, target location was specified by the assigned training sequence with a probability of .85 and by alternative sequence with a probability of .15. For example, if A was the assigned training sequence, then the bigram 3-1 was followed by a target at location 4 (from sequence A) with a probability of .85 and by location 2 (from sequence B) with a probability of .15. Locations from sequence A were referred to as probable targets while the ones from sequence B were called improbable targets. Shanks et al. (2005) also employed an alternative secondary task, the symbol-counting task, to exclude some of the nonintentional disruption effects associated with the tone-counting task (Stadler, 1995). As usual, the four target locations were arranged horizontally. One of four possible SRT target symbols (X, ?, $, and o) appeared on each trial. Single-task condition participants responded to the location of the target whereas dual-task condition participants were also required to count the combined number of X's and ?'s.

Following training, all participants completed Block 15 under single-task conditions with the same probable and improbable targets. Participants were then required to complete two generation tasks: an inclusion test in which the participants generated the training sequence and an exclusion test in which participants generated a sequence different from the training sequence. Each sequence of 100 digits was coded as 98 consecutive response triplets. The number of triplets that appeared in the assigned training sequence was calculated. For example, if the assigned training sequence was 242134123143 and the participant generated the sequence 2132, the triplet 213 appeared in the training sequence but the triplet 132 did not.

Shanks et al. (2005) computed the difference between RTs for probable and improbable targets in Block 15 for each participant. The difference was significantly greater in the single-task than the dual-task group reflecting better learning in the single-task conditions. Nevertheless, the mean difference was significantly greater than zero in the dual-task group. Thus, sequence learning was attenuated, but not eliminated, under dual-task conditions. It seems that participants obtained some explicit sequence knowledge since both groups could generate significantly more triplets that appeared in the assigned training sequence with inclusion than with exclusion instructions.

In summary, the studies conducted by Cohen et al. and Frensch et al. are difficult to interpret. Cohen et al. (1990) confounded task difficulty with sequence type. Frensch et al. (1998) used a within-subject design and could not determine whether the dual-task affected performance or learning. Unlike Frensch et al. (1998), Shanks and Channon (2002) used a between-subject design and found that dual-task training interfered with learning regardless of the testing conditions, replicating Nissen and Bullemer's (1987) finding. In a second study, Shanks et al. (2005) used an alternative secondary task and found attenuated sequential learning under dual-task conditions. Since no additional stimuli were involved, the adverse effect of the secondary task could be attributed only to the competition for limited attentional resource.

Stadler (1995) provided an alternative explanation of why sequential learning deteriorates as WM load increases. He disagreed with the notion that the secondary task hinders learning because participants need to withhold attention from the SRT task and claimed that the shifts of attention caused by the secondary task interfere with sequential learning. To test his hypothesis, Stadler used five conditions in a SRT task. In the tone-counting condition, an 83-ms tone, either high or low, was presented 17 ms following each response. Participants were required to count the number of high tones and write that number on a response sheet after each block of 42 trials. In the pauses condition, pauses between stimuli were varied by randomly mixing a longer response-to-stimulus interval of 2000 ms with the normal response-to-stimulus interval of 400 ms. The longer interval occurred at the same probability as the high tone in the tone-counting condition (i.e., 50%). Since there was no additional task for the participants to perform during the SRT task, this condition should not have increased working memory load. In the memory-load condition, a list of seven letters was presented before the SRT task for 5 seconds. Participants were instructed to remember those letters and report them immediately following the last of the 24 blocks of SRT trials. Since no secondary task was imposed during the SRT task, no disruption of sequential organization was expected. In addition to the three experimental conditions, two single-task control conditions were included. In the tones condition, participants were instructed to ignore the tones, either high or low, during the SRT task and in the letters condition, participants were told to ignore the letters presented before the SRT task. SRT learning was indexed by the difference between the mean of median RTs in Blocks 10-12 (repeating sequence) and in Blocks 13-15 (quasi-random sequence). The learning scores were highest in the control conditions, at an intermediate level in the memory-load condition, and lowest in the pauses and tone-counting conditions.

Because adding pauses during the SRT task would not create any additional attentional demand, and equivalent SRT learning scores were obtained in the pauses and tone-counting groups, Stadler (1995) argued that the tone-counting task does not compete for attentional capacity with the formation of sequential associations. Instead, he attributed the interference in both conditions to an organization problem: "... learning depends on practicing consistently organized runs of trials *(i.e., groups of successive trials)*, that shifts of attention may determine how the runs are organized, and the relation between attention and learning depends more on organization and intention than on capacity" (p. 674, Stadler, 1995). In contrast to Stadler's argument, it is possible that different mechanisms account for the performance decrements shown by participants in each of these groups. It may be that the response-to-stimulus interval of 2000 ms is too long for participants to form sequential associations automatically between successive stimuli whereas interference is produced when participants switched attention between the secondary task and primary task in the tone-counting condition.

Stadler (1995) also found that the performance of participants in the memory-load condition was poorer than that of participants in the control groups. Since no concurrent secondary task was imposed in the memory-load condition, voluntary verbal rehearsals might have interfered with SRT performance because participants switched attention between the primary and secondary tasks or because the secondary task competed for limited attentional capacity.

Stadler (1995) conducted an additional experiment using the memory-load condition and provided evidence that favored the attention-switching argument. Memory loads of five, seven, or nine letters did not differentially affect the rate of learning. It could be argued that since the stimulus onset intervals were constant, the frequency of switching attention between voluntary rehearsals and the primary implicit task was constant regardless of how many letters were to be remembered across different memory loads over the testing period. If so, level of memory-load would not be expected to correlate with implicit learning.

The Balls and Boxes Puzzle

The Balls and Boxes puzzle was first described by Kotovsky and Simon (1990) as a digital isomorph of the Chinese ring puzzle. ("Two problems are isomorphic if the graph of one problem can be mapped onto the graph of the other, with nodes and links corresponding one to one" p.147, Kotovsky & Simon, 1990.) For the traditional Chinese Ring puzzle (see Figure 1), the task is to remove five rings from a bar on which they are impaled. The device is three-dimensional; its parts are loosely joined, and can be twisted in many ways. One part can be slid over another in an effort to make a legal move. The Balls and Boxes puzzle, which was referred to as the low-information (Lo-Info) digital isomorph by Kotovsky and Simon (1990), is similar to the Chinese ring puzzle except that legal moves are shown in the display on the computer screen and participants move by positioning the mouse pointer over an open box and then clicking the left mouse button (see Figure 1).

The Balls and Boxes puzzle is a good candidate for investigating implicit learning. First, it is difficult to deduce the underlying rule structure of the puzzle from the initial description (Kotovsky, & Simon, 1990; Reber & Kotovsky, 1997). The linearity of the search space does not prevent most participants from making a large number of moves to reach a solution (Kotovsky, & Simon, 1990; Reber & Kotovsky, 1997). In other puzzles, such as Tower of Hanoi (Zanga et al., 2004), it is possible to plan long sequences of moves immediately and even deduce the optimal solution strategy from the instructions before working on the puzzle (Reber & Kotovsky, 1997; Zanga, et al., 2004). Second, as evidence against the possible contamination of explicit knowledge in SRT paradigm and other puzzles, Reber and Kotovsky (1997) found that no participant was able to provide a complete verbal description of the solution strategy. In their first experiment, verbal retrospective protocols obtained after the first solution trial were rated on a five-point scale. A score of one indicated that the protocol contained no useful information about the puzzle while a score of five reflected that a complete solution was described. No protocol earned a rating of 5. The majority of protocols (86.8%) were rated less than 3. It appeared that little conscious knowledge of the puzzle was acquired (Reber & Kotovsky, 1997). Third, trying to meet both information and sensitivity criteria defined by Shanks and St. John (1994), Reber and Kotovsky (1997) also used a move-selection test (Experiments 2 and 3), a strategy-statement questionnaire (Experiment 3), and the technique of concurrent verbal-protocol analysis (Experiment 4) to assess explicit knowledge about the puzzle. Across their four experiments, all participants failed these explicit tests and seemed to have learned strategies for solution implicitly.

The time needed to solve the Chinese Ring puzzle and its isomorphs has been shown to be sensitive to problem difficulty (Kotovsky & Simon, 1990) and working memory load (Reber & Kotovsky, 1997). Kotovsky and Simon (1990) examined the effect of varying working memory demands across the Chinese Ring puzzle and two isomorphic puzzles. The No-Info isomorph consists of a set of five boxes displayed on the screen. The Lo-Info isomorph (i.e., the Balls and Boxes puzzle) consists of a similar display except that information is provided as to which moves are legal by opening the corresponding boxes on the display (see Figure 1). Both the Chinese Ring puzzle and its two isomorphs followed the same basic solution rules. For the two isomorphs, the rightmost ball can always move, but the remaining four balls can be moved in or out of the boxes only when the ball immediately to the right is in its box and all other balls to the right are out of their boxes. The Chinese Ring puzzle is more difficult than the two isomorphs. None of Kotovsky and Simon's (1990) naive college participants were able to solve this puzzle within 2 hours. More time was needed to solve the No-Info isomorph (25.5 min) than the Lo-Info isomorph (14.6 min). However, when a hint was presented, 17 of 41 participants were able to solve the Chinese Ring puzzle within 2 hours. Average solution times decreased to 19.9 min for the No-Info isomorph and 13.0 min for Lo-Info isomorph (Kotovsky & Simon, 1990).

Reber and Kotovsky (1997) manipulated memory load by requiring participants to remember one, two, or three of the most recent letters they heard while learning the Balls and Boxes puzzle. The more letters they were instructed to remember, the greater was the load on working memory. They found that memory load correlated with the number of moves to solution on the first solution trial. However, on the second trial, working memory load did not correlate with performance. Thus, the secondary task affected learning but not subsequent transfer.

Across the implicit tasks studies (e.g., Cohen et al.,1990; Nissen & Bullener, 1987; Reber & Kotovsky, 1997; Shanks & Channon, 2002), only performance on the Balls and Boxes puzzle consistently deteriorated as working memory load increased. For this reason, the puzzle was used in the present study. Moreover, unlike sequence learning in which participants develop perceptual-motor associations (Shanks et al., 2005), the Balls and Boxes puzzle involves transforming material from an initial state to a specific goal state (Reber & Kotovsky, 1997) and might not depend on sequential learning. Finally, since the current state is wholly determined by previous moves, the choice made by participants between the two legal moves associated with 30 of the 32 states might reflect changes in their knowledge about the puzzle and provide researchers with useful indices.

The Speak-span task

Comparing individuals who differ in test-defined working memory capacity as a way of investigate the role of WM has not been studied using implicit learning tasks. An operational definition of working memory capacity might be the number of items that can be recalled in a memory task (Barrett et al., 2004). Individual-difference measures of working memory capacity (e.g., counting span task, operation span task, reading span task, and speak-span task) are impressive predictors of participants' performance on explicit learning tasks (Conway, Saults, & Elliott, 2002; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2002). Kane et al. (2001, 2002) argued that when attempting to learn complicated mental tasks, individuals with low WM capacity are less able to maintain all of the necessary information in working memory to construct a complex, integrated representation. The explicit system is assumed to be highly dependent on a limited-capacity WM system. Span scores have been shown to predict a variety of cognitive abilities, including rule-based learning (Smith & DeCoster, 2000), reading comprehension (Engle, Cantor, & Carullo, 1992), note taking (Kiewra & Benton, 1988, cited in Bleckley et al., 2003), and computer language learning (Shute, 1991, cited in Kane & Engle, 2003).

A variety of tasks have been used as measures of working memory capacity (Conway et al., 2005). One of these measures, the reading span task, was devised to measure both the storage and processing functions of working memory capacity for young adults (Howe & Rabinowitz, 1990). It has been frequently used and has been proven to be both reliable and predictive (Conway et al., 2005). Rabinowitz, Howe, and Saunders (2002) developed an age-appropriate modification for children, the speak-span task, as an index of working memory capacity. Using this task, they found a monotonic increase in scores on the task as a function of age and a positive correlation between speak-span scores and class-inclusion performance (Rabinowitz et al., 2002). Because the speak-span task is useful with 8 to 14 year-old children, it was employed to define children's working memory capacity in the present study.

Implicit Learning and Age

Unlike age-correlated changes in explicit learning, a relationship between age and the rate of implicit learning has been neither established (e.g., Schmitter-Edgecombe & Nissley, 2002; Vinter, & Perruchet, 2000) nor expected (Reber, 1993). The hypothesized neural basis for implicit learning (basal ganglia structures) is assumed to mature relatively early, whereas the prefrontal cortex systems, which are hypothesized to be involved in explicit learning, take longer to develop (Dienes, Broadbent, & Berry, 1991; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Reber, Gitelman, Parrish, & Mesulam, 2003; Seger, 1994). One might expect the rate of implicit learning to peak earlier in development than the rate of explicit learning (Huttenlocher & Dabholkar, 1997).

Karmiloff-Smith (1992) hypothesized that cognitive development consists of a succession of phases which are marked by different representational formats that change from implicit to explicit. The first phase achieved in each domain of competence corresponds to a level of behavioral mastery involving implicit knowledge. After more extended practice, those implicit representations will turn into explicit forms through an endogenous process of representational redescription (see Karmiloff-Smith, 1992 for details). Since the first phase is reached whenever a sufficient quantity of experience related to a task has been accumulated, implicit learning processes are age-insensitive. Consistent with her position, it has been demonstrated that many of the motor, perceptual, and cognitive acquisitions made by children in the course of development are implicit (e.g., Vinter & Perruchet, 2000; Gasparini, 2004; Reber, 1993).

Despite the fact that the age-invariant hypothesis is of theoretical interest, only a few researchers have compared samples of younger and older participants on implicit learning tasks. In most cases, no significant age differences have been found (e.g., Meulemans, Van der Linden, and Perruchet, 1998; Schmitter-Edgecombe & Nissley, 2002; Thomas & Nelson, 2001; Vinter & Perruchet, 2000). Meulemans et al. (1998) compared the performance of a group of children aged 6 to 7 years, a group aged 10 to 11 years, and a group of adults in a SRT task using a 10-trial sequence, 2413421431. A learning session was comprised of five blocks of 84 trials. Each block consisted of four random trials, followed by five presentations of a 16-trial sequence that started with the 10-trial repeating sequence and ended with six random trials. Following the learning session, participants were informed of the presence of a repeating sequence. Sixteen 4-stimulus sequences were constructed: eight sequences belonging to the repeating sequence (e.g., 2413, 4134, and 4214) and eight sequences which had never appeared in the fixed repeating sequence (e.g., 2412, 4131, and 4242). Participants were requested to rate these sequences on a five-point Likert scale: "I am sure I never saw it," "I believe I never saw it," "I don't know," "I believe I saw it," "I am sure I saw it."

Consistent with Reber's (1993) hypothesis that implicit learning is age-invariant, no age-related differences in the SRT performance were obtained. Learning appeared to have been implicit as participants did not discriminate the old and new recognition sequences (Meulemans et al., 1998).

Vinter and Perrucher (2000) provided additional evidence consistent with Reber's hypothesis of age-invariance. They selected the start-trace task to exploit a natural covariation present in drawing: the direction of movement in the tracing of closed geometrical figures is dependent on the starting point (i.e., the start-rotation rule). For example, if participants are presented with a starting point at 11 o'clock and required to trace over the circle, they will predominantly trace in a counterclockwise direction. Participants were instructed to trace the figures as fast and as accurately as possible during the training phase and were not informed about the purpose of the experimental manipulations (Vinter & Perrucher, 2000).

The experimental session comprised a familiarization phase, a training phase, a test phase, and a questionnaire phase. During the familiarization and training phases, two figures, a circle and a square, were chosen for tracing. A starting point was presented either at the 12 o'clock or the 6 o'clock position for the circle and at either middle point on the horizontals for the square. An arrow located 1 cm above or below the starting point was used to indicate the trace direction. Four types of training figures were presented: a top start with a counterclockwise trace (congruent), a bottom start with a clockwise trace (congruent), a top start with a clockwise trace (incongruent), and a bottom start with a counterclockwise trace (incongruent). Participants were randomly assigned to one of three training groups: the principle-incongruent group, the principle-congruent group, and the free-control group. The principle-congruent group received congruent instructions for 80% of the trials and incongruent instructions for the remaining 20% of the trials. Thus, the start-rotation principle was obeyed on 80% of the trials (i.e. 16 out of 20). The principle-incongruent group received a set of training figures organized in the same way as the principle-congruent group, except that the proportions of instructions that did or did not conform to the principle were reverted. Participants in the free-control group were asked to trace the figures without any information regarding starting location and movement direction.

After the training phase, all participants received two successive tests: a test in which only starting locations were presented as the cues followed by a test in which only trace directions were presented as the cues. Finally, participants filled in a free-report questionnaire and completed a cued-recall test in which they were given a set of figures and asked to state the starting point and the trace direction that they remembered having seen on the majority of trials during the training phase.

A comparison of 432 children between the ages of 4 and 10 years, and 54 adults failed to reveal any age differences in the way they implicitly learned principle-incongruent tracing behavior (Vinter & Perrucher, 2000). At all ages, principle-incongruent practice led to a large decrease in the production of drawing responses conforming to the start-trace principle when participants were asked to trace from a given starting position. The free-control and principle-congruent training groups did not differ in test performance. Participants from all groups were unable to describe any relevant information about the target manipulation in either the free reports or the cued-recall test, suggesting no explicit knowledge was obtained (Vinter

& Perrucher, 2000).

Schmitter-Edgecombe and Nissley (2002) investigated the age-invariance hypothesis using a matrix-scanning task, an implicit covariation learning task, with 72 older (age 59 to 83) and 72 younger adults (age 18 to 28). Participants first completed 30 practice trials in which they were required to search matrices for the quadrant location of the randomly located target "6". In the following five 60-trial blocks, participants were randomly assigned to one of three conditions. In the constant covariation condition (AAAAA), four matrices co-occurred with a unique location of the target throughout the scanning task. For example, when Matrix 1 was displayed, the target "6" always appeared in the upper left quadrant. The changed covariation condition was identical to the constant covariation condition with the exception that in the fourth block, the target appeared in the diagonally opposite quadrant of each matrix (AAABA). In the third condition, the no-covariation condition, the location of the target appeared in a random quadrant. Within each of the five training blocks, the target '6' appeared in each of the four quadrants 15 times.

Participants then received 64 test trials; 16 with each of the four matrices used in the training phase. They were informed that the presentation duration of the matrices would be so short that they probably would not be able to see the target '6'. They were also told that even though they might not detect the '6', they would still see it subliminally and make correct guesses. Following the test block, participants were presented with a implicit knowledge task in which they were asked to report how often they noticed the target '6' in the 64 trials on a six-point Likert scale, where '1' = never and '6' = very often. Finally, participants were asked for their general strategies in completing the task and whether they had noticed anything special about the matrix display.

The change in reaction time across the five training blocks differed significantly from zero for all groups except the older adult, no-covariation. The increase in reaction time from Block 3 to Block 4 in the changed covariation condition was significant ($p < .01$) for the older adults and approached significance ($p = .08$) for the younger adults. When the original pattern was reinstated, the reaction times of participants in the changed covariation condition decreased (Schmitter-Edgecombe & Nissley, 2002). In the explicit knowledge test, participants in all conditions exhibited chance performance. When asked whether they had noticed anything special about the displays, a larger percentage of older adults (86%) than younger adults (61%) reported no suspicions related to the matrices. However, none of the participants who acknowledged having been suspicious were correct in their assumption. In summary, although older adults were impaired relative to younger adults, both groups were able to implicitly process the manipulated covariation between matrices and target locations (Schmitter-Edgecombe & Nissley, 2002).

Occasionally, age-dependent implicit learning has been reported (Curran, 1997;

Howard & Howard, 1997; Maybery, Taylor& O'Brien-Malone 1995; Negash, Howard, Japikse, & Howard, 2003). Maybery et al. (1995) modified the incidental covariation task used by Lewicki (1986, as cited in Maybery et al., 1995). Children (5 to 7 and 10 to 12 years old) were trained with a number of 4×4 matrices of 16 pictures, divided into four quadrants. In the learning phase, children viewed several uncovered matrices in which positions of a particular picture, a house, correlated with two other stimulus features: the side from which the experimenter approached the child (left or right), and the color of the matrix board and cover (red or blue). During the test phase, the pictures were covered with a red or blue cloth, so that when the experimenter approached the child, both position and color cues were available. The children were required to guess the location of the house among the 16 covered locations. Finally, children responded to detailed questions concerning usage of cues in the guessing phase.

Implicit learning improved with age (Maybery et al., 1995). The 10-12-year-olds made more correct guesses in the test phase than did the 5-7-year-olds, whose performance was not above chance. In a subsidiary analysis, they found that the younger, as compared to older, children were less influenced by the covariation cues and more influenced by an uninformative cue, the house's previous position (Maybery et al., 1995). None of the children could spontaneously report any of the covariations in the initial open-ended questioning (Maybery et al., 1995).

Curran (1997) was interested in the relationship between sequence structure and implicit learning. He compared older adults (mean age = 67) and undergraduates using a SRT task with sequences that differed in pairwise transitions. Two types of sequences were used: a first-order predictive (FOP) sequence (121423431423), in which each location (e.g., 1) was followed by one location twice (4), another location once (2), and never followed by the third location (3); and a second-order predictive (SOP) sequence (121423413243), in which each of the possible bigrams (e.g., 42) uniquely determined the following item (e.g., 3). One would expect the FOP, as compared to the SOP, sequence to be easier to learn because performance on the FOP sequence would improve over training if the participants learned only 2-item transitions whereas 3-item transitions had to be learned with the SOP sequence. However, complete mastery of the FOP sequence required knowledge about up to 5-item transitions. For example, learning the different location followed by 21423 (4) and by 31423 (1) required knowledge of 5-item transitions which did not seem to be learned by younger adults (Curran, 1997). By contrast, the SOP sequence can be entirely learned with 3-item transitions. Therefore, the FOP sequence is actually more complex. Participants were trained for 9 blocks. Each block contained intermixed cycles of a 12 quasi-random trials (R) and a 12-item sequence (S) for a total 120 trials: RSSRSSRSSR. Following training, explicit knowledge was assessed using a five-point Likert scale and two sequence recognition tasks.

Younger participants learned both sequences equally well whereas older participants showed significant learning of only the SOP sequence. Smaller improvement over trials in SRT task was found for older, compared to younger adults. Neither younger nor older participants demonstrated explicit knowledge of the sequences (Curran, 1997).

Rather than a deficiency in the ability to learn subtle underlying patterns, age deficits in learning complex sequences might be due to the fact that older participants have difficulty in attending to targets that change spatial positions (Negash et al., 2003). To test this possibility, Negash, et al. (2003) compared 12 young (mean age = 20.67) and 12 older (mean age = 69.33) adults using a non-spatial variation of the alternating serial reaction time (ASRT) task in which random stimuli were embedded in a four-trial pattern (e.g., ArDrCrBr). In the non-spatial variant, target letters (A, B, C, or D) appeared in a box on the center of the screen. Because only one target location was used, neither a shift in visuospatial attention from one target location to another nor the use of eye movements was required. Participants were instructed to press the key corresponding to the target letter as quickly and accurately as possible. At the end of each block, participants were given feedback about their speed and accuracy on the preceding two blocks and told to focus more on accuracy when the accuracy scores were less than 86% or on speed if the accuracy scores were greater than 98%.

Participants completed five sessions, with 20 blocks of the ASRT task in each session. Each block began with 10 random trials which were followed by ten repetitions of an ASRT sequence. A questionnaire concerning strategy usage was presented at the end of each session. Following the five sessions, participants were given both recognition and preference tests. In the recognition task, participants were shown 20 randomly ordered trials. Each trial consisted of 16 items. The ASRT sequence used during the training was repeated twice in the 10 pattern trials whereas items were randomly determined in the 10 random trials. At the end of each trial, participants were asked to judge if these sequences occurred in the training on a four-point Likert scale. Similarly, in the preference task, participants were asked to rate how much they liked each of the sequences on a four-point Likert scale. People of both ages were able to learn the non-spatial sequence. However, neither group discriminated pattern and random sequences on the subsequent recognition and preference tests, suggesting the learning was implicit.

Negash, et al. (2003) classified errors on random trials in the ARST task into two categories: structure-consistent and structure-inconsistent errors. Suppose a participant, who had experienced the sequence ArBrCrDr, encountered a bigram AD that was followed by the stimulus D during training (that can only occur as part of the rDr sequence). If the participant incorrectly responded to the second D with a B (that can occur as part of the ArB sequence), the error would be considered to be a structure-consistent error, whereas a C response would be considered to be a structure-inconsistent error. Three aspects of the data reflected age-correlated deficits in implicit learning. First, older, as compared to younger, participants needed more practice to learn the sequence. Second, the difference of reaction times on pattern and random trials was smaller for older participants. Third, older participants generated a smaller proportion of structure-consistent errors on random trials and that proportion changed at a slower rate (Negash, et al., 2003).

In summary, inconsistent results have been reported about the relationship between implicit learning and age. There are a number of possible reasons that might account for the inconsistencies. First, age-correlated learning might reflect differences in the rate of explicit, rather than implicit learning (Maybery & O'Brien-Malone, 1998). The possibility that participants learn explicitly can not be totally eliminated (Shanks & St. John, 1994) and age-dependent explicit learning might have occurred in some studies but not others. Second, different implicit learning tasks may differ qualitatively in terms of the particular features learned and the neural substrates involved (Salthouse, McGuthry, & Hambrick, 1999; Seger, 1998). Similarly, even when the same task is used, features vary across stimulus manipulations and some may be associated with age-correlated difference in learning rate.

Finally, Seger (1994) identified three types of dependent measure – conceptual fluency, efficiency, and prediction and control – one or more of which had been used in implicit learning studies. Conceptual fluency measures tap participants' ability to rate or classify items as to whether they have the same structure as training stimuli. Participants usually report that they rely on their intuition or feelings in order to make these judgments (e.g., Meulemans et al., 1998). Efficiency measures are used to assess learning using speed or accuracy in responding to the stimuli (e.g., Curran, 1997; Negash et al., 2003; Schmitter-Edgecombe & Nissley, 2002). Prediction and control measures are used to assess participants' ability to predict accurately or to control some aspect of the stimuli (e.g., Maybery et al., 1995; Vinter & Perrucher, 2000). These classes of dependent measures were shown to tap different learning mechanisms associated with different knowledge representations (Seger, 1997). It is possible that different forms of representation are differentially correlated with age which may account for the diversity of age effects reported across implicit learning studies.

Objectives and Rationale

In the current study, the objectives were to determine whether implicit learning is affected by working memory capacity and/or age, and whether results obtained with adults in Reber and Kotovsky's (1997) study could be replicated with children. The speak-span task was selected to define participants' WM capacity. After each child completed the speak-span test, the Balls and Boxes puzzle was presented followed by a verbal report. The child then was asked to solve the puzzle a second time followed by a second verbal report. Finally, the child completed a move-selection test in which he or she was asked to choose the next move for each of the eight randomly selected states of the Balls and Boxes puzzle.

Working memory can be investigated either by comparing individuals who differ in test-defined WM capacity or by experimentally varying WM load using a secondary task. In general, increasing WM load and increasing WM capacity have reciprocal effects on task performance (Conway & Engle, 1996). Since it has been demonstrated that participants who have solved the Balls and Boxes puzzle can rarely describe anything about the structure of the problem and that the difficulty of the secondary task negatively correlates with learning rate (Kotovsky & Simon, 1990; Reber & Kotovsky, 1997), it was expected that higher speak-span scores would correlate with improved puzzle performance.

Fourth and sixth-graders (mean age ranges from 9.38 to 11.43 years old) were recruited as participants in the present study because there are marked age-related changes on many explicit tasks when children of this age range are compared (e.g., Bergling, 1999; Rebok, Smith, & Pascualvaca, 1997; Anderson, Nettelbeck, & Barlow, 1997). Since age correlates positively with WM capacity, if WM capacity influences implicit and explicit learning in a similar way, older children should learn the Balls and Boxes puzzle in fewer moves than younger children. Such a finding would be inconsistent with the Reber's (1993) hypothesis that implicit learning is age-invariant.

To the extent that implicit learning is age-invariant, it was expected that several results found with adults by Reber and Kotovsky's (1997) would be replicated with children in the present study. Performance would improve across trials; little knowledge of the puzzle would be revealed in their verbal retrospective protocols after each solution; and children would perform at the chance level on the move-selection test.

Method

Participants

Participants were 96 children attending a elementary school in the Eastern School District in the city of St. John's. There were 24 male and 24 female fourth graders and 24 male and 24 female sixth graders. The mean age of the fourth graders was 9.38 years (range = 8.83 to 10.42) and that of the sixth graders was 11.43 years (range = 9.25 to 12.50). Permission letters were sent to parents containing information about the study goals and procedures. Parents were asked to sign a consent form if they wanted their child to participate.

Apparatus and Stimuli

A computer was used to present the speak-span and the Balls and Boxes puzzle. The speak-span task was programmed in Quick Basic 4.5 and the Balls and Boxes puzzle was written in Microsoft Visual Studio .NET 2003. Stimuli were presented on a 12.1 inch widescreen display. Responses made during the speak-span task were entered by the experimenter using a keyboard. In the Balls and Boxes puzzle and move-selection test, the children responded using a two-button mouse.

The Speak-span Task

The speak-span task, which retained the structure but not the specific content of the reading-span task, was constructed using 110 English words ranging from 3 to 12 characters in length (see Table 1). The words were taken from fourth-grade books that were used in the St. John's school system, Newfoundland Board of Education. For each child, the words were randomly sorted into sets containing different numbers of words (2, 3, 4, 5, or 6). Five sets, all of which included the same number of words, made up a group. At the beginning of the speak-span task, children were provided instructions on the monitor (Rabinowitz, Howe, & Saunders, 2002, p.167):

> You will be presented with a number of sets of words. At the beginning of each set, you will be told how many words will occur in that particular set. I will then read you one word from the set. After I read the word, please make up a sentence using that word. After I type in your sentence, I will read the next word in the set. At the end of the set, you will be asked to remember each of the words that you made up a sentence about. You can remember the words in any order. After you have remembered as many words as you can, another set of words will start. The number of words per set will increase as the procedure continues. Before beginning the experiment, you will encounter a set of practice words.

The experimenter read these instructions to the child and answered any questions. Two-word sets were used in the pretest that ended as soon as the child correctly recalled both words in a set or a total of five two-word sets had been presented. After the pretest was completed, each child was presented with one to five test trials. A trial consisted of the presentation of five sets of words that constituted a group. Each set within a group contained the same number of words. Every child began with the two-word group. A child who reached criterion on the two-word trial was presented with a three-word trial and so forth. The most difficult trial involved six-word sets. At the beginning of each trial, the child was told how many words to expect in each set. After the experimenter read the first word aloud, the child was required to make up a sentence using that word. The experimenter typed the sentence as the child spoke. After all of the words in the set had been presented, the child was asked to recall the words in any order. Recall was self-paced. The experimenter typed the recalled words and hit the carriage return for each of the words in a set that the child omitted. Only the first three letters of the words that the experimenter entered were evaluated by the computer in an attempt to reduce the likelihood that typing errors could affect the outcomes. If the child successfully recalled the words in three of five sets making up a group, then the next trial was presented. Otherwise, the speak-span task stopped.

Speak-span scores range from 1 to 6. If a child was successful with fewer than two of the five sets in a group, then a speak-span score was assigned as the number of words in each set in the prior group. For example, if a child was successful with fewer than two sets on a 3-word trial, then a score of 2 was assigned. A speak-span score of 1 indicated that the child failed to succeed on at least two 2-word sets. If the child was successful in recalling exactly two of the five sets on a trial, then a speak-span score was assigned that was midway between the number of words per set on that trial and the previous trial (e.g., 2.5 on a 3-word trial). If the child was successful with more than two of the five sets, then they continued to the next level (unless they had completed level 6) (Howe, et al., 1998).

The Balls and Boxes Puzzle

As an isomorph of the Chinese Ring puzzle, initially each ball was in its box (see Figure 2, State 21). For each location in the problem space, there were never more than two operators that yielded legal moves (i.e., there were two boxes that were open indicating the balls that could be moved in or out of these boxes). The rightmost ball could always move; other balls could be moved if the ball immediately to the right was in its box and all other balls to the right were out of their boxes.

The problem space was fairly small with only 32 states. Figure 3 contains a complete description of the problem space, showing each possible state of the puzzle, numbering them by their distance from the goal state (State 0, the state where all the balls are out). The linear structure of the problem space is best appreciated by noticing that from any state except state 0 and 31, the only legal moves were to adjacent numbered states. State 0 and 31 were end states and only one move was associated with each of these states. Boxes opened and closed each time a move was selected. Given the small size of the problem space, 32 positions, it seemed surprising that it usually took as long as 5 to10 minutes to achieve a solution (Reber & Kotovsky, 1997). Exploration of the problem space was required to discover a strategy for solving the puzzle. Although this search was constrained because the problem space was linear, the direction of state change was not clear to the naive participants trying to solve the puzzle for the first time (Kotovsky & Simon, 1990).

The standard instructions appeared on the monitor:

You are going to solve the Balls and Boxes puzzle. The goal of the puzzle is to get five balls out of the five boxes. A ball can be moved into or out of its box using the computer mouse by clicking on the bar under the ball. A ball may only be moved into or out of its box if its box top is open. For instance, right now the two balls on the right could be moved but not the three on the left. If you make an illegal move, the computer will beep to remind you that is an error. As you move balls in and out of their boxes, the box tops will open and close. The trick to the puzzle is to move the balls to get the correct boxes to open up so that you can move all the balls out of their boxes.

The experimenter read these instructions to the child and answered any questions. In the task, the underlying rule that determined whether a ball could be moved was not mentioned to the children. In the initial state presented, State 21, all five balls were in boxes. In the goal state (i.e., state 0), all of the balls were out of the boxes.

Retrospective Report

Each time the children completed the Balls and Boxes puzzle, they were asked to describe the strategies they had used to solve the puzzle and to attempt to describe how the puzzle worked. Since a free recall test would reduce the likelihood that low confidence knowledge was reported, children were informed that it was better to report information that might be wrong rather than to omit information that might be true (Shanks & St. John, 1994). The protocols were rated on a five-point scale. A protocol was given a rating of 1 if it contained no informative statements about the puzzle (e.g., containing only statements such as "there were five balls and five boxes"). A rating of 2 indicated that the protocol contains one somewhat informative statement (e.g., "The leftmost ball was hard to get out.") For a protocol to receive a rating of 3, it had to include several somewhat informative statements (e.g., "The ones on the left were hard. In order to get these out, you had to keep taking out and replacing the ones on the right.") A rating of 4 or better was assigned if the protocol contained partial rule information that could significantly aid solution (e.g., "The leftmost ball could be removed when the right three were out, the second from the left could be removed when the right two were out") or contained a general strategy (e.g., "Whenever a box opens, move the ball, otherwise move the rightmost ball.") A rating of 5 was assigned to those who could describe a complete solution.

Two people rated 20 verbal protocols, 10 for each trial. They disagreed on five of these protocols, but the rating difference was all one. One rater gave the children higher scores on three of the five discrepant protocols. The correlation of the ratings was .765, $p < .001$, even though the scores only ranged from one to three.

Move-selection Test

After the second free recall trial was completed, children were given a move-selection test to determine if they could choose correct moves when problem states were presented in isolation. Each participant was presented 8 randomly selected isolated states from the puzzle and asked to select the move to get one step closer to the goal state.

Procedure

Each child was tested individually in a small room for approximately 35 minutes. When the child entered the room, the investigator asked the child's name, birth date and grade. The child then was informed that he/she could withdraw from the study at any time if he/she wished to do so. The speak-span task was then presented and followed by the Balls and Boxes puzzle. The delay between these tasks was determined by the time to load the Balls and Boxes software. After solving the puzzle, participants were asked to talk about "how you solved the puzzle, how the puzzle worked and especially anything you could say that would help someone else solve the puzzle." A tape recorder was used to record the child's description. The puzzle was then presented a second time. Finally, following a second request for a description of the solution, a move-selection test was presented to assess whether child could choose correct moves when problem states were presented in isolation.

Results

In order to provide an initial description of the data using all the independent and dependent variables of interest, a linear correlation matrix and factor analysis were computed. These were followed by a series of analyses of variance and regression analyses, conducted to better understand the relationships among variables. Finally, comparisons were made between the child data and the adult data reported by Reber and Kotovsky's (1997). The independent variables were gender, age, and trial. The dependent variables included children's speak-span scores and the number of correct responses on the move-selection test. All the remaining dependent variables were assessed on Trial 1 and Trial 2: total moves to solve the puzzle (Move 1 and 2), solution time (Time 1 and 2), the number of moves that violate one of the problem rules (Illegal 1 and 2), the number of reversals produced by clicking on the same ball on consecutive moves (Reversal 1 and 2), the number of error-free moves made immediately prior to reaching the goal state (called final-path length in Reber and Kotovsky's study, 1997) (Final 1 and 2), and ratings of the verbal protocol obtained after each problem solution (Verbal 1 and 2). Note that the overall moves to solve the puzzle measures included illegal moves, reversal moves, and error-free final moves on the same trial.

Linear correlation matrix

As can be seen in Table 2, age correlated positively with speak-span scores, negatively with solution time (p = .050 for Time 1 and p = .098 for Time 2), and positively with both verbal protocol measures (p = .003 for Verbal 1 and p =.009 for Verbal 2). The ranges of speak-span scores and the relationship between age and speak-span scores are consistent with data reported by Rabinowitz et al., (2002). In addition, the negative correlation between age and solution time is consistent with data reported by Kail (1984, 1993). Surprisingly, children's verbal protocol ratings on both trials increased with age suggesting some explicit learning occurred. However, speak-span scores did not correlate with any variable except age. Thus performance on the Balls and Boxes puzzle appears to be independent of test-defined WM capacity. With the exception of final-path length, the problem performance measures correlated within but not across trials.

Factor Analysis

A principal component factor analysis was conducted in which all the independent and dependent variables were components. Based on the correlations described above, one would expect that Trial 1 and Trial 2 performance variables, except for final-path length, would load on independent factors. Since verbal ratings on Trials 1 and 2 correlated with all measures on Trial 2 except final-path length, one would also expect both verbal protocol ratings would load on the Trial 2 factor.

The first three factors accounted for 55.41% of the variance in the data matrix. Factor 1, 2 and 3 accounted for 25.27%, 20.11%, and 10.04% of the variance respectively. As can be seen in Table 3, all Trial 1 performance measures, with the exception of the final-path length, loaded on the second factor, named Trial 1 Factor. Similarly, both verbal protocols and all Trial 2 performance measures except for the final-path length loaded on the first factor, named Trial 2 Factor. Since speak-span scores were not correlated with children's verbal protocols on either trial. It was surprising that speak-span and both verbal protocol measures, along with age, loaded on the third factor, named Explicit Learning Factor. To the extent that the third factor does reflect individual differences in explicit learning, the Speak-span test appears to have been a sensitive index of children's WM capacity.

It is clear from the results in the correlation matrix and factor analysis that the hypotheses involving WM and age were not supported. Inconsistent with the hypothesis that higher speak-span scores would correlate with improved puzzle performance, children's speak-span scores did not correlate with any variable except age and did not load on either of the Trial factors. Similarly, inconsistent with the hypothesis that older children would learn the Balls and Boxes puzzle in fewer moves than younger children, age did not correlate with any performance measure except solution time and did not load on either of the Trial factors.

Analyses of Variance

In all the analyses of variance, gender and grade were the between-subjects factors and if appropriate, trials was the within-subject factor. The dependent variables were the performance measures, verbal protocol ratings, and the move-selection scores. Across all the analyses, the only contrast effects that reached significance ($p < .05$) were main effects of age and trials (see Table 4 for Means, Standard Deviation, F-values, and p-values). Speak-span scores and ratings of the verbal protocols increased with age while time to solution decreased with age. However, except for the age effect on solution time, neither WM capacity nor age influenced children's performance. Therefore, the hypotheses that higher speak-span scores would correlate with improved puzzle performance and older children would learn the Balls and Boxes puzzle in fewer moves than younger children were not supported. Three replication hypotheses were supported. First, performance improved from the first to the second trial on all measures. Second, only two children were perfect on their second trial performance. Third, as was the case with adults in the Reber and Kotovsky's (1997) study, children performed at the chance level on the move-selection tests (see Table 4).

Regression Analyses

In the regression analyses, temporal sequence was used in selecting the predictor variables. Measures collected before a particular variable was assessed were used as predictors. For example, Age, Gender, Speak Span, Time 1, Move 1, Final 1, Illegal 1, and Reversal 1 were used to predict Verbal 1 ratings. If multiple R^2 was significantly greater than chance, the influence of each predictor variable was assessed with all the other predictor variables partialled out. The regression analyses are summarized in Table 5.

On Trial 1, none of multiple R^2s associated with the performance measures (Time 1, Move 1, Final 1, Illegal 1, and Reversal 1) was significant. Thus, consistent with the conclusions draw based on linear correlation matrix, the factor analysis, and the analysis of variance, neither WM capacity nor age influenced children's performance on their first solution. However, the multiple R^2 associated with the verbal ratings on Trial 1 was significant as were the related partial correlation coefficients associated with Gender, Move 1, Final 1, and Illegal 1 (see Table 5). Boys were more likely to verbalize useful knowledge about the puzzle than girls. The larger number of the overall moves, the longer the final-path length and the fewer the number of illegal moves, the higher the verbal protocol ratings were on Trial 1. The interpretation of these significant partial correlations is not clear.

The distribution of Trial 1 verbal protocols is summarized in Table 6. Note that 43.8% of the children failed to provide any relevant information, another 43.8% could describe only one somewhat informative piece of information (e.g., "The leftmost ball was hard to get out" or "The rightmost box was always open"), and 12.5% of the children offered more than one informative statement (a rating of 3) or partial rule (a rating of 4). Therefore, similar to the data obtained with adults in Reber and Kotovsky's (1997) study, little knowledge of the puzzle was revealed in children verbal retrospective protocols after each solution.

Although most children failed to provide much information about how to solve the problem, the first protocols were predictive of their second protocol ratings and all Trial 2 performance except for the final-path length on that trial (see Table 2 and Table 5). Inspection of Table 7, the difference in Trial 2 performance of children with protocol ratings of 1 and 2 was particularly striking. It appears that minimal explicit knowledge was an important aid to problem solving on Trial 2.

Comparison between Children and Adults from Reber and Kotovsky's (1997) study

Because similar procedures were employed [a], data obtained with children in the present study were compared with data with undergraduates in Reber and Kotovsky's (1997) Experiments 1 and 2. Inspection of Table 8 reveals substantial developmental change on several of the dependent variables. For instance, the undergraduates' mean final-path lengths were 18.0 (SD = 2.8) and 17.3 (SD = 3.0) moves on Trials 1 and 2, respectively, approaching the perfect score of 21 moves between the starting and goal states. In contrast, the children's mean final-path lengths were about half the length of the undergraduates, 7.9 and 10.1 moves on Trial 1 and 2, respectively. The average difference between the final-path length of the adults and children on Trial 1, 10.1 moves, was greater than 3 times the standard deviation of the adult mean. Similarly, the discrepancy on Trial 2, 7.2 moves, was almost 2.5 times the standard deviation of the adult mean on that trial. Therefore, less sequential knowledge was obtained by children. Moreover, on average, the children needed 59 more moves to solve the puzzle than did the adults on Trial 1. This difference is greater than 4 times the standard deviation of the adult mean. Similarly, the discrepancy in the number of moves made by children and adults in solving the puzzle on Trial 2 was 53, approximately 6 times the standard deviation of the adult mean on that trial. Therefore, the adults learned the puzzle with greater efficiency than did the children and the probability that the children and adults were sampled from the same population is minute.

Since the adults in Reber and Kotovsky's (1997) Experiments 1 and 2 were required to describe how they solved the puzzle only following their first solution, only Trial 1 protocol ratings were compared. As can be seen in Table 9, the distributions of the ratings of the Trial 1 verbal protocols were similar: 87.6% of the children and 86.8% of the adults obtained a rating less than 3, χ^2 (3) = 4.97, $p > .05$. Therefore, the developmental change in the percentage of participants who acquired explicit information about the puzzle is unlikely to account for the substantial differences in performance between children and adults.

Discussion

Learning

The improved performance on the second trial replicates Reber and Kotovsky's (1997) results and shows that children acquired knowledge about the puzzle when solving it the first time. Among the 10 children who solved the puzzle in the fewest moves (less than 24 moves) on Trial 2, one received a rating of 4, two a rating of 3, five a rating of 2, and two a rating of 1 on the first trial verbal protocol. None of these children described the basic rule (i.e., except for the rightmost ball, which can always move, a ball can be moved only when the ball immediately to the right is in its box and all other balls to the right are out of their boxes), and only one child's protocol would have been useful to a naive problem solver. Therefore, the statements given by these children reflect, at best, partial knowledge of the Balls and Boxes Puzzle. However, to show improvement in solving the puzzle, explicit knowledge is not necessary.

There are four types of information that could be learned about the puzzle: (a) the basic rule, (b) avoiding illegal moves, (c) avoiding reversal moves, and (d) the move sequence. First, if one knows the basic rule for the puzzle and the order to get the balls out, then it is possible to deduce the next correct move. None of the children described the basic rule and only five children referred to move order.

Second, since illegal moves do not facilitate problem solution, elimination of such moves will reduce both time and moves to solution. As can be seen from the decrease of the proportion of illegal moves across trials (i.e., from 7.9% to 4.0%), children learned to avoid such moves either explicitly or implicitly. Although only one girl referred to illegal moves in her protocols, it is likely that most children explicitly acquired information about such moves from the instructions and did not consider illegal moves to be worth mentioning. On the other hand, some children may have implicitly learned to avoid illegal moves when they were occasionally reminded by the beeps produced by the computer following this type of error. In order to get a legal ball in or out of a box, children had to move the computer mouse to the corresponding bar under that box and click the left mouse button. Because the five boxes were connected and the children were clicking on the bar rapidly (see Table 10), the ballistic quality of the moves made it difficult to stop an illicit move once it had been initiated. After practice, however, children may have implicitly learned to reduce "ballistic" errors.

Third, a participant who avoids reversals would achieve a reasonably efficient solution. If one chooses randomly at the start state and never reverses the last move, progress is guaranteed to either the goal state in 21 moves or to the top of the problem space, State 31, in 10 moves. A participant who directly moves to State 31 and then makes only the forced reversal would reach the goal state in a total of 41 moves. The substantial decrease of reversal moves across trials (i.e., from 37.5 to 21.6) shows that children acquired some knowledge about avoiding reversals. However, only 7 and 11 children mentioned this information in their protocols on Trials 1 and 2, respectively. Thus, most children were likely to have learned to avoid reversals implicitly.

Fourth, children's inability to differentiate correct and incorrect moves on the move-selection test indicates that cues associated with prior moves are used to solve the problem. The length of the final-path increased significantly across trials (i.e., from 7.91 to 10.08) and the number of children with final-path lengths greater than 20 on Trial 2 was triple the number on Trial 1 (i.e., 4 vs. 12). Thus, patterns of move sequence probably were learned and used to aid Trial 2 performance. The sequence of moves used to solve the puzzle errorlessly is highly patterned. For example, starting at

State 31, the move sequence 545 3 545 2 545 3 545 1 545 3 545 2 545 3 545 (spacing added for emphasis) ends at the goal. Therefore, some children were likely to learn part or all of the sequence explicitly. For instance, one sixth-grade boy who solved the puzzle in 150 moves with a protocol rating of 2 on Trial 1 improved his performance to 21 moves on Trial 2 and described most of the sequence following his errorless performance. Moreover, the first verbal reports of two boys, a fourth- and a sixth-grader, were rated 4 and both boys appeared to have learned the puzzle explicitly. The fourth grader solved the puzzle in 23 moves on Trial 2 and described most of the move sequence in his Trial 1 protocol providing enough information to help a naive participant solve the problem. The sixth grader, who solved the puzzle in 25 moves on Trial 2, described parts of the move sequence and the strategy of moving the rightmost two balls to get the others out of the boxes before the second solution.

Explicit Learning

It has been argued that verbal reports are thought to be insensitive measures of explicit learning (Perruchet & Amorin, 1992; Shanks & St. John, 1994) because they do not tap low confidence knowledge (Berry & Dienes, 1993). In addition, it is possible that children misinterpret free recall instructions as requests to report only rules. Nevertheless, differences in verbal protocol ratings on Trial 1 were predictive of Trial 2 performance, replicating the result Reber and Kotovsky (1997) obtained with adults.

Explicit learning might have occurred during the instructions, during the course of solving the puzzle on Trial 1, and when children were required to verbalize relevant information about the puzzle. For example, consider "avoid-illegal moves." During the instructions, because illegal moves were explicitly demonstrated while the experimenter was reading the instruction to avoid such errors, most children were likely to gain relevant information and remember it throughout the experiment. During problem solving, when a child made an illegal move, the computer beeped and presumably increased the probability of explicit learning and recall on Trial 1. Finally, verbalization may change the way people use knowledge acquired earlier (Lane & Schooler, 2004). As a result, before making a verbal report, many of the children might not have been aware of the "avoid-illegal moves" knowledge. It is possible that at the moment they were required to verbalize what had been learned, children began to examine relevant information available in short-term memory and generated explicit knowledge about avoiding illegal moves and other features of the puzzle.

Implicit Learning

Implicit learning is characterized by behavioral sensitivity to the structure of the environment and the lack of awareness of this sensitivity (Berry & Dienes, 1993; Cleeremans, 1993; Cleeremans, 1997; Kirkhart, 2001; Reber, 1993). It is distinguished from learning that occurs with concurrent awareness (explicit learning). Reber and Kotovsky (1997) based their conclusions that implicit learning occurred with the Balls and Boxes puzzle on four findings. First, participants were not able to specify how they had solved the puzzle in the written retrospective protocols after the first solution (Experiment 1 and 2). Second, participants performed at chance level on the move-selection test (Experiments 2 and 3). Third, almost all participants made errors when solving the puzzle on Trial 2. Fourth, when participants were required to describe their thoughts about choosing which balls to move during problem solving, reportable strategies did not develop (Experiment 4).

Move-selection test, verbal reports, and Trial 2 performance data were collected in the present experiment and Reber and Kotovsky's (1997) findings were replicated. Similar to the undergraduates in Experiment 2 (Reber & Kotovsky, 1997), children failed to recognize correct moves for isolated states in the move-selection test, and presumably, did not acquire explicit knowledge about single moves.

With the exception of the two boys who received a Trial 1 protocol rating of 4, children were unable to describe how they had solved the puzzle following Trial 1. More than half the children obtained a protocol rating of 2 or 3 on that trial which indicates that one or two features of the problem space were correctly identified. However, these features had no functional value for a naive problem solver. The remaining children (43.8%) did not provide any relevant information in their Trial 1 protocols, although their performance also improved across trials. In short, children learned to solve the puzzle and improved across trials even though most of their verbal protocols contained no useful information about how to solve the puzzle. Thus the obtained knowledge is unverbalized and presumably implicit.

Similar to the undergraduates in Reber and Kotovsky's study (1997), only a small number of children performed perfectly on Trial 2. Therefore, the knowledge obtained during the first solution, whether implicit or explicit, was not complete. Since most participants explored the problem space again and were unable to describe useful information about solving the puzzle following Trial 1, Reber and Kotovsky (1997) argued that the errorless moves at the end of the first trial did not reflect explicit knowledge and therefore, the errors that occurred on Trial 2 indicated that the knowledge acquired on Trial 1 was incomplete and implicit.

In summary, if the move-selection test, verbal reports, and errorless Trial 2 performance are appropriate indices of explicit learning, then the chance level performance in the move-selection test, the low verbal protocol ratings, and the Trial 2 errors showed that most children learned the Balls and Boxes puzzle either completely or partially implicitly.

Can WM Capacity Influence Implicit Learning?

Berry and Dienes (1993) and Reber (1993) suggested that implicit learning is relatively insensitive to the availability of attentional resources. Nissen and Bullemer (1987) first investigated this conjecture by comparing learning under single- and dual-task conditions in the SRT task and showed that the secondary task had an adverse effect on sequence learning. Several investigators followed up their findings (e.g., Cohen et al., 1990; Frensch et al., 1998; Shanks & Channon, 2002; Shanks et al., 2005) and found that performing a secondary task seems to interfere with implicit learning in the SRT task.

Learning on the Balls and Boxes puzzle also consistently deteriorated as WM load increased, associated with a secondary task (Kotovsky & Simon, 1990; Reber & Kotovsky, 1997). Reber and Kotovsky (1997) used a secondary task in both learning and transfer stages. Consistent with Shanks and Channon's result (2002), the imposition of the dual task interfered with learning but did not affect transfer. In the present study, in order to examine whether test-defined WM capacity also would affect implicit learning, the performance of children who differed in speak-span scores were compared on the Balls and Boxes puzzle. Surprisingly, the wide range of speak-span scores (i.e., from 1 to 6) did not correlate with performance measures on either the learning or transfer trial.

Stadler (1995) provided an explanation that might account for the different effects of test-defined and manipulated working memory on implicit learning. He assumed that if the participants attended to the sequence, associations between successive stimuli are formed, strengthened, and stored automatically. Cowan, Saults, and Ellion (2002) also claimed that sequential associations form each time the participant focuses on the parts of the sequence appearing on the screen. Stadler (1995) distinguished two dimensions of attention: the control of attention (i.e., the switching of attention among different tasks) and the capacity of attention (i.e., number of tasks that can be simultaneously attended to). He claimed that the shifts of attention, not capacity, caused by the secondary task interfere with sequential organization. Thus, without any distractions, participants attend to the sequence, form sequential associations automatically, and do not need to rely on WM. In contrast, when a secondary task is added, participants shift their attention between tasks which interferes with the development of sequential associations in the primary task.

Reber and Kotovsky's (1997) required participants to listen to an audiotape that consisted of a stream of letters and to remember the most recently presented one, two, or three letters. Stadler (1995) would argue that the dual-task causes shifts of attention between forming sequential associations in the primary task and updating the most recent letter in the concurrent secondary task. Since the letters appeared at the same rate across the three memory-load groups, disruption of the sequential organization would be equivalent across conditions. However, solution time increased with memory load, suggesting that simple task switching cannot account for the data. Consistent with Shanks et al's (2005) findings, attentional capacity was shown to be necessary to perform the secondary task.

In the present study, no secondary task was used. Children's speak-span scores did not correlate with performance measures or verbal protocol ratings. Therefore, if the speak-span task is an appropriate assessment of children's WM capacity, it appears that in the absence of a secondary task, WM does not affect learning or transfer in the Balls and Boxes puzzle. It appears that sequential associations form automatically when participants attend to the stimuli in implicit learning tasks.

Can Age Influence Implicit Learning?

The hypothesized cognitive hardware required for implicit processing (i.e., basal ganglia structures) is available at birth and predates the development of the structures required for conscious awareness (i.e., prefrontal cortex system) which gradually becomes available developmentally (Reber, Gitelman, Parrish, & Mesulam, 2003; Seger, 1994; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Dienes, Broadbent, & Berry, 1991). Therefore, Reber (1992, 1993) claimed implicit learning should be insensitive to developmental changes in contrast to age-dependent explicit learning (Siegler, 1998).

As can be seen in Table 4, data consistent with Reber's age-independent assumption were obtained in the current study. With the exception of time to solution and the verbal protocol ratings, no significant differences in performance as a function of age were found on either trial. The substantial differences in time to solution were not surprising. Kail (1991, 1993) reviewed speed of processing studies on speeded tasks and found that response time negatively correlated with age during childhood. Inspection of Table 6 reveals that although the mean verbal protocol ratings of sixth graders was higher than fourth graders on both trials, only one sixth-grader and one fourth-grader obtained a rating of 4, a value which reflects useful explicit information. Thus, the acquisition of useful explicit information was probably age-invariant.

Besides the exception of time to solution and the verbal protocol ratings, the remaining performance measures indicate that implicit learning occurred at about the same rate in both grades. This age-invariant result is consistent with findings reported in other studies in which the implicit learning of 4 to 11 year-old children was compared (e.g., Healy, 2003; Meulemans & Van der Linden, 1998; Thomas & Nelson, 2001).

In contrast, when data obtained with children in the current study were compared with that of the undergraduates in Reber and Kotovksy's (1997) Experiments 1 and 2, substantial age differences were found. Each of the following three hypotheses may account for the age-invariant data obtained from fourth- and sixth-graders and the age-dependent results obtained by comparing the children in the present study with the adults in Reber et al.'s (1997) study. First, the age difference between forth and sixth graders might be too constrained. It may be necessary to sample a wider age range to find developmental changes in implicit learning. Second, it may be that age-invariant implicit learning reflects that the relevant knowledge base is stable across the age range studied. Age-dependent implicit learning may be dependent on changes in the relevant knowledge base (see Murphy, McKone, & Slee, 2003). Third, the rate of implicit learning might be stage-dependent, mirroring Piaget's stages in cognitive development. If so, evidence consistent with Reber's age-invariance hypothesis would only be obtained in constrained periods of development.

Conclusions

Data obtained with children in the present study replicated several of the key findings obtained by Reber and Kotovsky (1997) with college students. Performance of both children and adults improve across trials on the Balls and Boxes puzzle. Most of what they learn is implicit as little useful information appears in verbal protocols, performance is at chance level on the move-selection test, and errors are made when they resolve the puzzle.

Since the speak-span test, an explicit index, did not correlate with performance measures on Trials 1 and 2 in the Balls and Boxes puzzle, it appears that individual differences in WM are unrelated to implicit learning in this task. This finding is consistent with Cowan et al.'s (2002) assumption that sequential associations usually form automatically when participants attend to the stimuli in implicit tasks.

Consistent with the data reported in implicit learning studies involving children from 4 to 11 yeas of age (e.g., Healy, 2003; Meulemans & Van Linden, 1998; Thomas & Nelson, 2001), implicit learning occurred at about the same rate in fourth- and sixth-graders in the present study. However, substantial developmental differences in implicit learning appear when the performance of younger adults is compared to that of middle-age children on the Balls and Boxes puzzle. If these striking age differences can be replicated in other implicit tasks, then Reber's age-invariance hypothesis would only be applicable in constrained periods of development.

References

Anderson, M., Nettelbeck, T., & Barlow, J. (1997). Reaction time measures of speed of processing: Speed of response selection increases with age but speed of stimulus categorization does not. *British Journal of Developmental Psychology, 15* (2), 145-157.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130* (4), 553–573.

Bergling, B. M. (1999). Developmental changes in reasoning strategies: Equating scales for two age groups. *European Psychologist, 4* (3), 157-164.

Berry, D. C. (1997). (Ed.), *How implicit is implicit learning?* New York: Oxford University Press.

Berry, D. C., & Dienes, Z. (1993). *Implicit Learning: Theoretical and empirical issues.* Hove, U. K.: Erlbaum.

Bleckley, M. K., Durso, F., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review, 10* (4), 884-889.

Cleeremens, A. (1993). *Mechanisms of implicit learning: A connectionist model of sequence processing.* Cambridge, MA: MIT Press.

Cleeremans, A. (1997). Principles for implicit learning. In D. C. Berry (Ed.), *How*

Implicit is Implicit Learning? (pp.195-234). New York: Oxford University Press.

Cleeremens, A., & Jimenez, L. (1998). Implicit sequence learning: The truth is in the details. In M. A. Stadler, & P. A. Frensch, (Eds), *Handbook of implicit learning* (pp. 323-364). Thousand Oaks, CA: Sage.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 17-30.

Cowan, N., Saults, J. S., & Elliott, E. M. (2002). The search for what is fundamental in the development of working memory. *Advances in Child Development and Behavior, 29*, 1-48.

Conway, A. R. A., & Engle, R. W. (1996). Individual difference in working memory capacity: More evidence for a general capacity theory. *Memory, 4* (6), 557-590.

Conway, A. R. A., Kane, M. J., Buting, M. E., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12* (5), 769-786.

Currarn, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological Research, 60*, 24-41.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 875-887.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18* (5), 972-992.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition, 22*, 95-110.

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 47-104). Thousand Oaks, CA: Sage.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, 12 (1), 13-18.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: the effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychology Research, 61*, 83-98.

Gasparini, S. (2004). Implicit versus explicit learning: Some implications for L2 teaching. *European Journal of Psychology of Education*, XIX, 203-219.

Hambrick, D. Z, & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: an investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*, 339–387.

Healy, R. M. (2003). Implicit and Explicit learning of Serial Reaction Time Task.

Unpublished master's thesis. Memorial University of Newfoundland, NL.

Howard, J. H., & Howard, D.V. (1997). Age differences in implicit learning of higher order dependences in serial patterns. *Psychology and Aging, 12*, 634-656.

Howe, M. L., & Rabinowitz, F. M. (1990). Resource panacea? Or just another day in the developmental; forest. *Developmental Review, 10*. 125-154.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology, 387* (2), 167-178.

Kail, R., & Nippold, M. A. (1984). Unconstrained retrieval from semantic memory. *Child Development, 55*, 944-951.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin,109* (3), 490-501.

Kail, R. (1993). The role of a global mechanism in developmental change in speed of processing. In M. L. Howe & Pasnak (Eds), *Emerging themes in cognitive* development. Volume I: Foundations (pp. 97-119). New York: Springer-Verlag.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130* (2), 169-183.

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 & Review 9* (4), 637-671.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132* (1), 47-70.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science. Cambridge*, MA: MIT Press.

Keele, S., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychology Review, 110*, 316-339.

Kirkhart, M. W. (2001). The nature of declarative and nondeclarative knowledge for implicit and explicit learning. *The Journal of General Psychology, 128* (4), 447-461.

Kotovsky, K., & Simon, H. A. (1990). Why are some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology, 22*, 143-183.

Lane, S. M. & Schooler, J. W. (2004). Skimming the Surface: Verbal Overshadowing of Analogical Retrieval. *Psychological Science, 15* (11), 715-719.

Maybery, M., & O'Brien-Malone, A. (1998). Implicit and automatic processes in cognitive development. In K. Kirsner, C. Spee;man, M. Maybery, A.

O'Brien-Malone, M. Anderson, & C. Macleod (Eds.), *Implicit and explicit mental process.* Mahway, NJ: Erlbaum.

Maybery, M., Taylor, M., & O'Brien-Malone, A. (1995). Implicit learning: Sensitive to age but not IQ. *Australian Journal of Psychology, 47*, 8-17.

Meulemans, T., & Van der Linden, M. (1998). Implicit sequence learning in children. *Journal of experimental child psychology, 69*, 199-221.

Murphy, K., McKone, E., & Slee, J. (2003). Dissociations between implicit and explicit memory in children: The role of strategic processing and the knowledge base. Journal of Experimental Child Psychology, 84, 124-165.

Musen, G., & Squire, L. R. (1993). Implicit learning of color-word associations using a Stroop paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1083-1100.

Negash, S., Howard, D. V., Japikse, K. C., & Howard, J. H. (2003). Age-related differences in implicit learning of non-spatial sequential patterns. *Ageing Neuropsychology and Cognition, 10* (2), 108-121.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32.

Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *Journal of Experimental Psychology, 134* (3), 368–387.

Perruchet, P., Chambaron, S., & Ferrel-Chapus, C. (2003). Learning from implicit learning literature: Comment on Shea, Wulf, Whitacre, & Park (2001). *The Quarterly Journal of Experimental Psychology, 56A* (5), 769-778.

Rabinowitz, F. M., Howe, M. L., & Saunders, K. (2002). Age, memory load, and individual differences in working memory as determinants of class-inclusion reasoning. *Journal of Experimental Child Psychology, 81*(2), 157-193.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118* (3), 219-235.

Reber, A. S. (1992). The cognitive unconscious: An evolutionary perspective. *Consciousness and Cognition, 1*, 93-133.

Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious. New York: Oxford University Press.

Reber, A. S., & Kotovsky, K. (1997). Implicit learning in problem solving: The role of working memory capacity. *Journal of Experimental Psychology: General, 126* (2), 178-203.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating Explicit and Implicit Category Knowledge with fMRI. *Journal of Cognitive Neuroscience, 15* (4), 574-583.

Rebok, G. W., Smith, C. B., & Pascualvaca, D. M. (1997). Developmental changes in attentional performance in urban children from eight to thirteen years. *Child*

Neuropsychology, 3 (1), 28-46.

Salthouse, T. A, McGuthry, K. E., & Hambrick, D. Z. (1999). A framework for analyzing and interpreting differential aging patterns: Application to three measures of implicit learning. *Aging, Neuropsychology, and Cognition, 6* (1), 1-18.

Seger, C. A. (1994). Implicit learning. *Psychology Bulletin, 115* (2), 163-196.

Seger, C. A. (1997). Two forms of sequential implicit learning. *Consciousness and Cognition, 6*, 108-131.

Seger, C. A. (1998). Multiple forms of implicit learning. In M. A. Stadler & P. A. French (Eds.), *Handbook of implicit learning* (pp. 295-322). Thousand Oaks, CA: Sage Publications.

Schmitter-Edgecombe, M., & Nissley, H. M. (2002). Effects of aging on implicit covariation learning. *Aging Neuropsychology and Cognition, 9* (1), 61-75.

Shanks, D. R., & Channon, S. (2002). Effects of a secondary task on "implicit" sequence learning: learning of performance? *Psychological Research, 66*, 99-109.

Shanks, D. R., Rowland, L. A., & Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological Research, 69*, 369–382.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17* (3), 367-447.

Siegler, R. S. (1998). *Children's thinking (3rd ed.).* New Jersey: Prentice Hall.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4* (2), 108-131.

Stadler, M. A. (1995). Role of attention in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21* (3), 674-685.

Thomas, K. M., & Nelson, C. A. (2001). Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology, 79,* 364-387.

Vinter, A., & Perruchet, P. (2000). Implicit learning in children is not related to age: Evidence from drawing behavior. *Child Developmeny, 71* (5), 1223-1240.

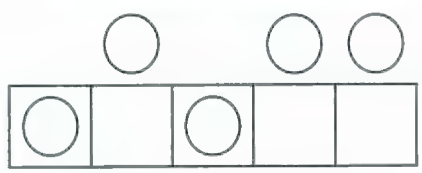
Zanga, A., Richard, J., & Tijus, C. (2004). Implicit learning in rule induction and problem solving. *Thinking and Reasoning, 10* (1), 55-83.

Footnote

[a] There were four WM load groups in Reber and Kotovsky's (1997) Experiments 1 and 2. Participants in the no load condition were not required to perform a secondary task and experienced the same procedure as did the children in the current study. Since WM load interfered with learning in the three load groups, the differences between the performance of the children in the present study and adults in Reber and Kotovsky's (1997) study was reduced by combining the data of the no load and load groups.

(a) Chinese Ring Puzzle

(b) No-Info digital isomorph



(c) Lo-Info digital isomorph

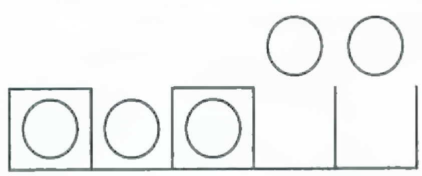


Figure 1. The Chinese Ring puzzle (a) and two digital isomorphs, No-Info (b) and Lo-Info (c).

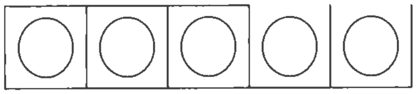


Figure 2. Initial appearance of the Balls and Boxes Puzzle.

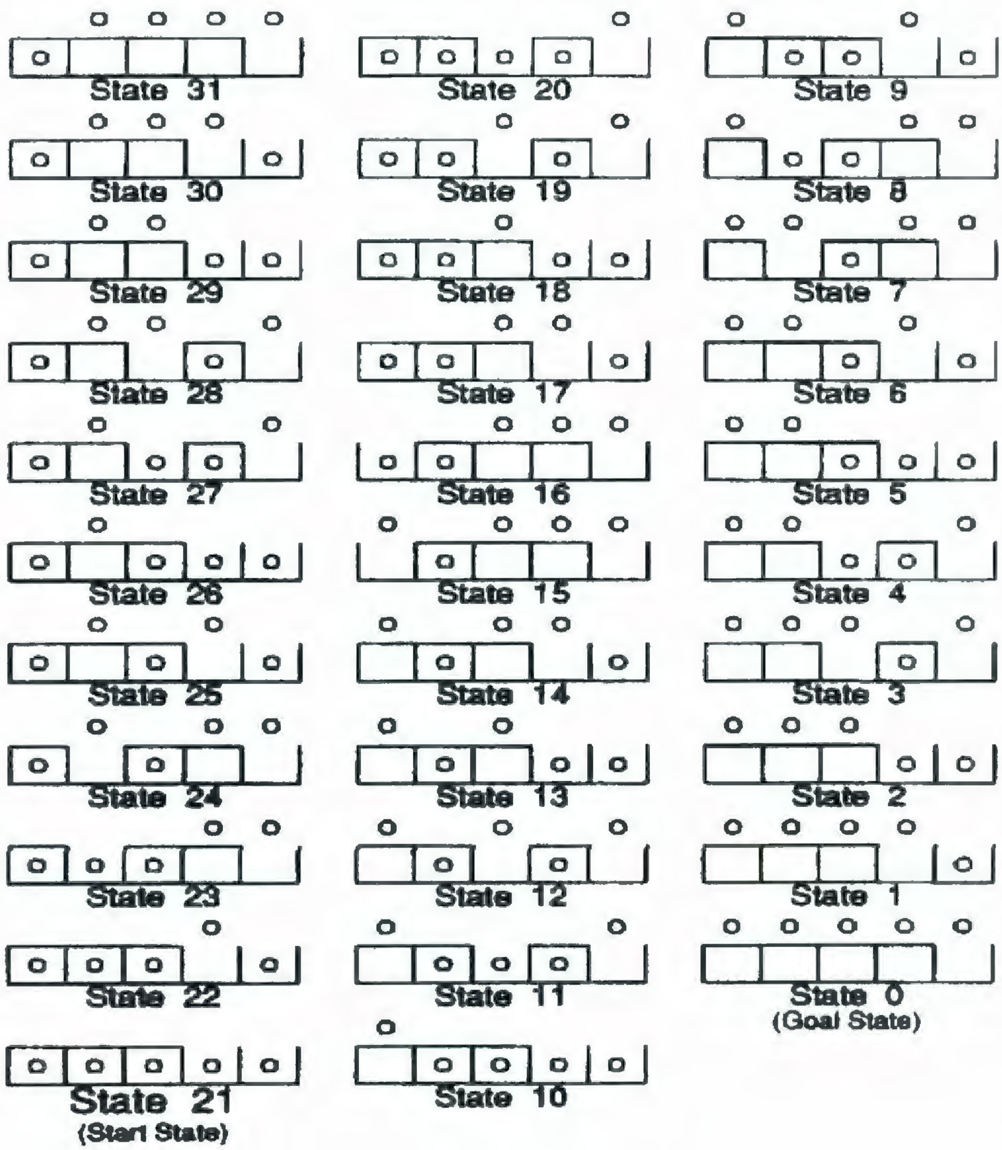


Figure 3. The problem space of the Balls and Boxes puzzle.

Each state is shown with the state number underneath. The problem space is linear, so that states with consecutive numbers are connected in the problem space (e.g., from State 21, it is possible to move to States 22 or 20 but no others).

Table 1. Words Presented in the Speak-Span Task

adult	Africa	airplanes	alcoholic	animal
art	astronaut	basketball	beach	blind
body	Canada	carpenter	cavemen	centimeter
chair	change	chipmunks	complain	cone
cub	diamonds	different	dinosaurs	dollar
driftwood	earth	eat	elephant	excuse
exercise	famous	fat	feeling	flowers
food	forest	fossil	geese	gills
glass	gold	grown-ups	healthy	hibernate
history	hobby	hockey	hole	hot
hungry	hunting	iron	jungle	kitten
large	learn	listening	lungs	medicine
messy	migration	mistake	money	Mount-Everest
mouse	museum	music	Monday	nest
noise	Ontario	painting	park	picture
planet	plants	poison	Prime-Minister	province
sad	school	season	secret	serious
shiver	silence	size	snake	snowflakes
spider	stars	steal	stomach	stories
storm	strange	summer	safe	team
temper	tiny	Toronto	tough	wallet
whales	winter	worm	year	zoo

Table 2. Correlation Matrix

	Gender	Months	SP	T1	M1	I1	R1	F1	V1	T2	M2	I2	R2	F2	V2	MS
Gender	1.000															
Months	-.072	1.000														
SP	.007	.232*	1.000													
T1	.098	-.201*	-.084	1.000												
M1	.044	.038	.060	.832**	1.000											
I1	-.028	.105	.012	.521**	.715**	1.000										
R1	.036	-.104	-.001	.693**	.903**	.732**	1.000									
F1	.145	-.013	-.011	-.012	-.012	-.205*	-.108	1.000								
V1	-.170	.304**	.173	-.084	.022	-.208*	.077	.275**	1.000							
T2	.076	-.170	-.074	.097	.027	.020	.000	-.091	-.394**	1.000						
M2	.013	.114	.022	.007	.052	.098	.080	-.092	-.330**	.889**	1.000					
I2	-.073	-.134	.006	-.082	-.050	.190	.004	-.219*	-.227*	.594**	.686**	1.000				
R2	.076	.130	.071	.034	.023	.139	.093	-.154	-.383**	.805**	.881**	.531**	1.000			
F2	.053	.046	-.009	.306**	.154	-.099	.009	.188	.001	-.011	-.149	-.253*	-.225*	1.000		
V2	.125	.266*	.154	-.114	.013	-.204	.044	.234*	.613**	-.479*	.388*	-.306**	-.409	.065	1.000	
MS	.045	.067	.143	.172	.158	.047	.147	.026	.158	-.030	-.045	-.153	.036	.032	.027	1.000

Note. SP=Speak Span Task, T 1=Time 1, T 2=Time 2, M 1=Move 1, M 2=Move 2, V 1=Verbal 1, V 2=Verbal 2, MS=Move-Selection, I 1=Illegal 1, I 2=Illegal 2, R 1 = Reversal 1, R 2 = Reversal 2, and F 1=Final 1, F 2= Final 2. * $p < .05$; ** $p < .01$; *** $p < .001$ (2-tailed).

Table 3. Component Matrix for the Factor Analysis

	Component		
	Trial 1 Index	Trial 2 Index	Explicit Learning Index
Gender	.053	.082	-.242
Month	-.006	-.320	.518
Speak-span Score	.019	-.088	.587
Time 1	.837	.241	-.133
Time 2	-.263	.833	.113
Move 1	.936	.228	.140
Move 2	-.243	.850	.308
Illegal 1	.711	.406	.059
Illegal 2	-.247	.688	.270
Reversal 1	.869	.312	.113
Reversal 2	-.230	.826	.285
Final 1	-.056	-.302	-.085
Final 2	.228	-.180	-.262
Verbal 1	.081	-.630	.535
Verbal 2	.151	-.673	.424
Move-selection Test	.222	-.089	.291

Table 4. Means, Standard Deviations, F-values, and p-values associated with the Main Effects of Grade and Trials

	Grade 4	Grade 6	F (1, 92)	*p*	Trial 1	Trial 2	F (1, 92)	*p*
No. of total moves	135.49 (106.01)	121.06 (89.58)	1.05	NS	153.83 (99.80)	102.72 (89.93)	14.28	<.001
Time to solution	299.02 (231.40)	218.81 (188.59)	7.75	<.01	338.39 (225.47)	179.45 (169.59)	32.60	<.001
Final-path	8.82 (6.04)	9.17 (6.41)	0.13	NS	7.91 (5.57)	10.08 (6.66)	7.25	<.01
Verbal protocol	1.66 (0.77)	2.19 (0.87)	14.36	<.001	1.71 (0.74)	2.14 (0.92)	30.56	<.001
No. of illegal moves	10.65 (16.78)	6.96 (7.65)	3.47	NS	12.31 (13.58)	5.29 (11.73)	17.57	<.001
No. of reversals	32.99 (34.79)	26.07 (24.10)	2.54	NS	37.45 (35.62)	21.61 (20.49)	15.03	<.001
Speak-Span	2.78 (0.53)	3.19 (0.95)	6.60	<.05				
Move-Selection	49.3% (17.2%)	48.8% (18.7%)	0.02	NS				

Table 5. Regression Analyses

Predicted Variables	Predictors	R^2	df	F	p	Significant partial correlation coefficients
Time 1	Set 1	0.049	(3, 92)	1.58	0.200	N/A
Move 1	Set 1	0.008	(3, 92)	0.25	0.863	N/A
Final 1	Set 1	0.021	(3, 92)	0.66	0.577	N/A
Illegal 1	Set 1	0.014	(3, 92)	0.42	0.736	N/A
Reversal 1	Set 1	0.012	(3, 92)	0.38	0.768	N/A
Verbal 1	Set 3	0.294	(8, 87)	4.52	0.000	Gender*, Move 1*, Final 1*, Illegal 1*
Time 2	Set 2	0.180	(9, 86)	2.10	0.038	Verbal 1**
Move 2	Set 2	0.142	(9, 86)	1.58	0.136	N/A
Final 2	Set 2	0.247	(9, 86)	3.13	0.003	Time 1*
Illegal 2	Set 2	0.154	(9, 86)	1.74	0.092	N/A
Reversal 2	Set 2	0.196	(9, 86)	2.33	0.021	Verbal 1**
Verbal 2	Set 2	0.418	(9, 86)	6.87	0.000	Verbal 1***
Move-Selection	Set 2	0.098	(9, 86)	1.04	0.414	N/A

Set 1: Gender, Age, and Speak Span; Set 2: Set 1 plus Verbal 1, Time 1, Move 1, Final 1, Illegal 1, and Reversal 1;

Set 3: Set 1 plus Time 1, Move 1, Final 1, Illegal 1, and Reversal 1.

* $p < .05$; ** $p < .01$; *** $p < .001$ (2-tailed).

Table 6. Frequency and Percentage of Different Levels of Protocol Rating for both Grades on both Trials.

	Verbal 1				Verbal 2			
Protocol Rating	Grade 4	Grade 6	Total	Percent	Grade 4	Grade 6	Total	Percent
1	29	13	42	43.8	19	6	25	26.0
2	16	26	42	43.8	19	22	41	42.7
3	2	8	10	10.4	9	17	26	27.1
4	1	1	2	2.1	1	3	4	4.2
5	0	0	0	0	0	0	0	0

Note. Protocol ratings range 1-5. Higher numbers indicate a protocol with more information about how to solve the Balls and Boxes puzzle.

Table 7. Retrospective Protocol Ratings and Corresponding Means and SD of Second Trial Performance

Protocol rating after Trial 1	No. of children	Move2	Time2	Illegal 2	Reversal 2	Verbal 2
1	42	137 (111)	261 (207)	8.7 (16.8)	30.9 (23.9)	1.5 (0.7)
2	42	79 (60)	121 (97)	2.6 (3.2)	15.2 (14.2)	2.4 (0.5)
3	10	71 (49)	113 (93)	3.2 (4.4)	13.5 (12.3)	3.1 (0.3)
4	2	24 (1)	25 (2)	0 (0)	2 (0)	4 (0)

Table 8. Means and Standard Deviations of obtained Scores from Children in the present Study and Undergraduate Adults in Reber and Kotovsky's (1997) Study.

Group	No. of children	Move 1	Time 1 (s)	Move 2	Time 2 (s)	Final 1	Final 2	Move-Selection
Children	96	154 (100)	339 (225)	103 (90)	179 (170)	7.9 (5.6)	10.1 (6.7)	49% (17%)
Adults	14	94 (12)	192 (39)	49 (9)	82 (23)	18.0 (2.8)	17.3 (3.0)	55% (3.3%)

Table 9. Frequency and Percentage of Different Levels of Protocol Rating for Children and Adults on Trial 1

Protocol Rating	Children [a]	Percent	Adults [b]	Percent [c]
1	42	43.8	43	56.6
2	42	43.8	23	30.3
3	10	10.4	6	7.9
4	2	2.1	4	5.2
5	0	0	0	0

Note:

a. Averaged across grades;

b. Experiment 1 of Reber and Kotovsky (1997), compile across memory load conditions;

c. Since the meaning of the rating of 1.5, 2.5, 3.5, and 4.5 was not specified in Reber and Kotovsky's paper, the protocol ratings of 1 and 1.5 were combined to a rating of 1 and etc.

Table 10. Mean time on Each Move for Grade 4 and Grade 6 on both Trials

	Trial 1 (s)	Trial 2 (s)
Grade 4	2.44	1.89
Grade 6	1.95	1.57