JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 56, 328–349 (1993)

# Chess Expertise and Memory for Chess Positions in Children and Adults

WOLFGANG SCHNEIDER

*University of Wuerzburg, FRG*

HANS GRUBER

*University of Muenchen, FRG*

ANDREAS GOLD

*University of Frankfurt, FRG*

AND

KLAUS OPWIS

*University of Freiburg, FRG*

This paper presents a replication and extension of Chi's (1978) classic study on chess expertise. A major outcome of Chi's research was that although adult novices had a better memory span than child experts, the children showed better memory for chess positions than the adults. The major goal of this study was to explore the effects of the following task characteristics on memory performance: (1) Familiarity with the constellation of chess pieces (i.e., meaningful versus random positions) and (2) familiarity with both the geometrical structure of the board and the form and color of chess pieces. The tasks presented to the four groups of subjects (i.e, child experts and novices, adult experts and novices) included memory for meaningful and random chess positions as well as memory for the location of wooden pieces of different forms on a board geometrically structured by circles, triangles, rhombuses, etc. (control task 1). Further, a digit span memory task was given (control task 2). The major assumption was that the superiority of experts

We thank Gabi Golling and Markus Brunner for assistance in collecting and analyzing data. Thanks are also due to David Bjorklund, Chuck Brainerd, Bob Siegler, and two anonymous reviewers for comments and valuable suggestions on earlier drafts of the manuscript. Reprint requests should be sent to Wolfgang Schneider, Institut fuer Psychologie, University of Würzburg, Wittelsbacherplatz 1, D-97074 Würzburg, Germany.



0022-0965/93 $5.00

should be greatest for the meaningful chess positions, somewhat reduced but still significant for the random positions, and nonsignificant for the board control task. Only age effects were expected for the digit span task. The results conformed to this pattern, showing that each type of knowledge contributed to the experts' superior memory span for chess positions. 

During the past 2 decades, numerous studies have demonstrated the importance of the knowledge base for memory performance (for reviews see Chi & Ceci, 1987; Schneider & Pressley, 1989). Content knowledge seems to be one of the crucial sources of memory development in childhood and adolescence, often outweighing other relevant factors like capacity, strategies, and metamemory (cf. Anderson, 1990; Bjorklund, 1990; Siegler, 1991).

Some of the most impressive findings concerning the impact of the knowledge base on memory performance stem from research that compared experts and novices in the domain of chess (cf. Chase & Simon, 1973; de Groot, 1946, 1978). Chess is a useful domain for the study of skilled performance (cf. Ericsson & Crutcher, 1990). There are several million players worldwide, and chess ratings that provide an independent measure of playing strength are often available. Compared to many other domains, it is thus easy to define expertise in chess.

Analyses of chess experts' and novices' memory performance have revealed both quantitative and qualitative differences. The chess board reconstruction task used by Chase and Simon (1973) seems particularly suited to illustrating these differences. In this task, subjects are required to reconstruct from memory chess positions that have been presented for only a short time. Chase and Simon demonstrated that compared to novices, their expert subject (a chess master) recalled larger sequences of chess pieces. Recall was in rapid bursts separated by noticeable pauses. Chase and Simon suggested that performance on the chess board reconstruction task depended on the ability to encode the chess positions in "chunks" (i.e., configurations of pieces). They concluded that differences in the complexity of preexisting patterns or chunks in long-term memory accounted for differences in recall of chess players of different levels of skill.

From a developmental perspective, the major advantage of the expert/novice paradigm is that knowledge and chronological age are not necessarily confounded: It is not only possible to recruit adult chess novices but also to find child chess experts in elementary school for experimental studies. The classic developmental study was conducted by Chi (1978), who recruited experienced and inexperienced chess players and gave them the chess board reconstruction task described above. The most interesting aspect of this research was that subjects' knowledge correlated negatively with age: The six children (average age = 10 years) were the experts;

the six adults were the novices. Chi found that the children's short-term memory for chess positions was superior to that of the adults. Moreover, Chi replicated Chase and Simon's (1973) findings that the child experts recalled more chunks and more pieces per chunk than the adult novices. On the other hand, this advantage was restricted to the domain of expertise. The typical adult superiority in short-term memory capacity was present, thus demonstrating that the children's superiority on the chess task was not due to any general processing advantage. Due to small sample size, however, the differences between children's and adults' memory span were not statistically significant. Chi concluded that short-term memory capacity was not inherently a function of age, but rather of content knowledge.

Chi's research attracted much attention among developmental psychologists. As is true with many studies that seem intuitively convincing, close replication experiments were considered unnecessary even though the data base was small (see Schneider & Hasselhorn, in press, for a more detailed discussion of this problem). From a methodological point of view, both the small sample size of Chi's study and the fact that only two of the four possible groups (i.e., child and adult experts and novices) were included call for a validation of results. To our knowledge, however, there have been only two developmental studies on chess expertise related to Chi's research.

In one of these studies, Roth (1983) compared child and adult experts and novices on a chess board comparison task. Subjects had to determine if pairs of boards were the same or different. Results concerning the speed of the comparison process for meaningful chess boards were contrasted with two independent controls: randomly arranged pieces on chess boards, and randomly arranged digits on chess boards. Roth found that differences in knowledge accounted for the observed adult/child variance. Knowledge of the stimulus domain enhanced processing rates in both children and adults to about the same extent. The magnitude of the knowledge effect was sufficient to eliminate any significant differences between child and adult experts. Further, the knowledge effect accounted for between-age group differences in that child experts outperformed adult novices. Thus Roth's findings on speed of perceptual comparisons paralleled Chi's results on short-term memory processes.

A second study on child expertise was conducted by Horgan and Morgan (1990). Horgan and Morgan focused on true child chess experts, that is, school-age children who played competitive chess and who had official chess ratings. Their sample of 113 child experts included a fairly large proportion of the top child chess players in the United States. Two experiments conducted by Horgan and Morgan were replications and extensions of the classic memory and board reconstruction studies of de

Groot (1946, 1978) and Chase and Simon (1973) described above. As a main result, it was found that performance on the original chess board reconstruction task was significantly related to chronological age and chess rating: older and higher-rated players performed better on the memory task. When a modified version of the board reconstruction task (copying task) was given, where subjects did not need to rely on memory but were allowed to look back at the target board, the number of glances was negatively correlated with chess ratings. The better players used their chess knowledge to organize information more efficiently, as indicated by larger perceptual chunks. These findings indicate that individual differences in chess knowledge are significantly correlated with performance on the board reconstruction task. This result seems particularly remarkable given the high expertise of the entire child expert sample under study.

Although the studies by Roth (1983) and Horgan and Morgan (1990) seem to confirm the basic conclusions of Chi's research, they cannot be considered close replications. The most apparent differences concerned task materials and sample characteristics. As noted above, Roth did not focus on memory but rather perceptual processes. Although Horgan and Morgan did include memory tasks in their study, their board reconstruction tasks were not directly comparable to the one used by Chi (1978).

Furthermore, definitions of chess expertise varied across studies. There is not much information on the chess skill of Chi's (1978) sample. From the six child experts recruited from a local chess tournament, only one had an official chess rating. The adult novices were able to play chess "to some degree" (Chi, 1978, p. 81). Roth's (1983) descriptions of expertise were at least as imprecise. The only information on child experts is that they had placed first in local chess tournaments. According to Roth, adult "experts" had some experience playing chess but they were not accomplished players. The fact that child and adult novices had no experience in chess at all makes it difficult to compare the findings for the novice groups with the results of Chi's study, where the novices knew how to play chess. Horgan and Morgan (1990) recruited true child chess experts for their experiments. However, as expertise and age were not systematically varied in their study, effects of expertise and age could not be independently assessed.

This short summary of developmental studies of chess expertise indicates that Chi's classic research has not been replicated in detail. Thus one goal of the present study was to provide a close replication of Chi's study with a larger sample. Another major goal of our study was to extend the original design in at least two aspects: First, like Roth (1983), we included subgroups of child novices and adult experts in our sample. From a methodological point of view, both knowledge (expert/novice) and age (child/adult) have to be manipulated, a priori, to yield unconfounded estimates of the impact of each factor on memory performance. An additional advantage of such a design is that child and adult experts' and novices' strategies observed during task solution can be directly compared.

A second extension concerned a procedure aimed at identifying possible sources of the experts' superior memory performance. As noted by Ornstein and Naus (1985), one general problem with studies of expertise is that an association between expert status in a particular area and differential patterns of recall of this material does not constitute an explanation of how such differences arise. The critical issue is to determine how experts are able to use their better structured knowledge in the service of remembering.

In our view, several aspects of the chess board reconstruction task influence experts' superior memory performance: First, greater familiarity with the configurations of chess pieces on the board, and greater knowledge of their meanings enables experts to represent the relevant information in larger chunks. Greater familiarity with the characteristics of the chess board (i.e., geometrical pattern, form and color of chess pieces) seems to be a second source of experts' superiority. If this second type of information is influential, experts should outperform novices even when chunking strategies seem no longer effective, that is, when the task is to reconstruct random chess positions. Although this assumption does not square with Chase and Simon's (1973) observation that recall of random positions was comparable for their three subjects (one master, one good chess player, one novice), results of subsequent studies on this issue (e.g., Holding & Reynolds, 1982; Lories, 1987; Reynolds, 1982) showed that chess expertise was positively correlated with recall of random chess positions. For example, Lories (1987) demonstrated that adult chess experts recalled more random positions than chess novices after a 60-s study period. Although the difference between experts and novices was greater for meaningful chess positions than for random positions, comparisons on both tasks yielded statistically significant results.

Since we do not believe that chess experts have better spatial abilities than chess novices, we expected no performance differences on a control task that required the reconstruction of wooden pieces on a board that had little in common with a chess board. Performance on this task should also not be affected by chronological age, because children are very good at memory reconstruction tasks using visual stimuli (cf. Schneider & Pressley, 1989). Yound children's great performance in the Memory game may serve as an example for this phenomenon. If our assumption is correct, the knowledge-related advantages of experts should be eliminated in this task setting.

We used three board reconstruction tasks to test our assumptions. Two meaningful chess positions, similar to those used by Chi (1978), were presented first. Like Chi, we assessed immediate recall and repeated recall of chess positions. In the immediate recall task, subjects immediately placed the appropriate chess pieces on a blank board (trial 1). On the repeated recall task, up to four more repetition trials were given, or until perfect performance was attained. Improvement across trials was considered an index for learning ability. In addition, a delayed recall task was given to explore aspects of long-term memory. Without previous warning, subjects were asked to reconstruct the first meaningful chess position from memory after having completed the trials on the meaningful chess position. The expectation was that large differences between experts and novices should be found for immediate recall, improvement across trials, and delayed recall, regardless of age.

We also presented one random chess position. Again, immediate recall and improvement across trials were assessed. Although we still expected significant expert/novice recall differences, they were expected to be less substantial than those found for the meaningful chess positions (see Lories, 1987).

In a third step, the board control task was given. Subjects had to reconstruct the positions of wooden pieces (cylinder, cone, sphere, prism, cube, cuboid) of two different colors on a board geometrically structured by circles, triangles, and rhombuses. Again, immediate recall and improvement across trials were assessed. No significant differences among the four groups were expected for this task.

Finally, a WISC digit span task was provided as a second control assessment. Here, the expectation was that adults would outperform children, regardless of chess expertise.

A short interview followed at the very end of the session. Subjects were asked to indicate how often (per week) they practiced and played chess and when they started with the game. Information about their school performance also was obtained.

## METHOD

*Subjects*

A total of 40 children and 40 adults participated in the study. The mean ages for the children and adults were 11.9 years (range: 10.0–13.4) and 26.8 years (range: 22.0–42.0), respectively. The groups of children and adults were further divided (20 each) on the basis of chess knowledge. All child experts were active members of Bavarian chess clubs. Most of them had participated in Bavarian chess championships. Although in Germany no official chess ratings are available for child players, results from various championships showed that our child experts were among the best Bavarian players in their age group. All child novices knew about the rules of the chess game and had played occasionally for a short time (less than 8 months). There were only boys in the child sample. Although not all of the adult experts were members of chess clubs, all had played on a regular basis for more than 10 years. All adult novices knew about the chess rules but had played only occasionally for less than 5 years. The adult experts and novices (22 male, 18 female) were students and faculty members of a German university. In order to validate our classification of experts and novices, the Knight's Tour task described below was given to all subjects.

*Materials and Procedure*

All subjects were tested individually. The experimental tasks were presented in a single session lasting about 60 min. Each subject was given five tasks: two chess reconstruction tasks (meaningful and random chess positions), a board control task, a digit span task, and the Knight's Tour task. For half of the subjects, the two meaningful chess positions were presented first, followed by the random position, the board control task, and the digit span task. For the other half of the subjects, the two control board positions were given first, followed by the random position and the two meaningful chess positions. The Knight's Tour task was always presented at the end of the session. All sessions with children were videotaped.

*Chess board reconstruction tasks.* All subjects were shown chess boards containing 22 chess pieces. Two *meaningful positions* were taken from German chess magazines. These two positions were used repeatedly for all subjects. The positions came from games played by master players, which, however, were not known to any of the subjects. One *random position* was generated from the meaningful positions by placing the same pieces on a chess board but in a random arrangement (see Fig. 1 for examples of a meaningful and the random chess position).

In the meaningful chess board reconstruction task, subjects were told that a chess board would be presented for 10 s and that their task would be to reconstruct this constellation of chess pieces on an empty board as accurately as possible. After a practice trial, the meaningful chess board reconstruction task was given with two different positions. Immediate recall was assessed first, followed by up to four repetition trials, separately for each position. Before each trial, the board was emptied, and the subjects restudied the respective position for 10 s. Reconstruction began immediately after the target arrangement had been removed. If a subject managed to completely reconstruct the chess board before trial 5, no more repetition trials were given. Several of the experts but none of the novices succeeded before trial 5. Delayed recall was also assessed for the first meaningful chess reconstruction task. That is, subjects were asked to reconstruct the first meaningful position after they had gone through repeated trials of the second meaningful chess board reconstruction task.

The procedure used for the random chess board reconstruction task

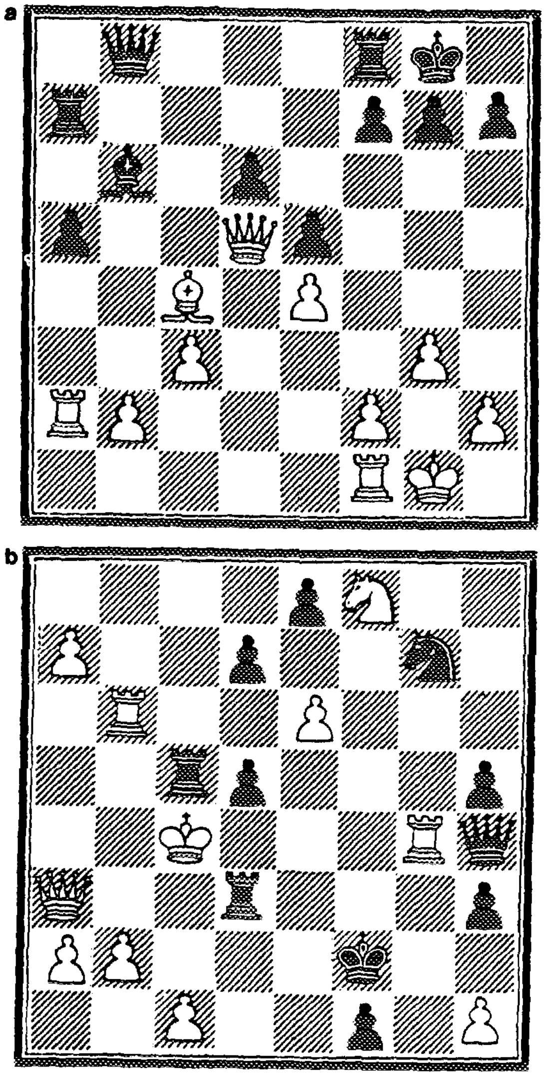


FIG. 1. Examples of (a) meaningful and (b) random chess positions presented to the subjects.

was identical to the one described for meaningful chess patterns, with the exceptions that only one position was provided across the five trials because of time restrictions and that no delayed recall condition was established. The number of chess pieces correctly reconstructed during the first trial (i.e., immediate recall) and the number recalled on the following four trials were used as dependent variables for this task.

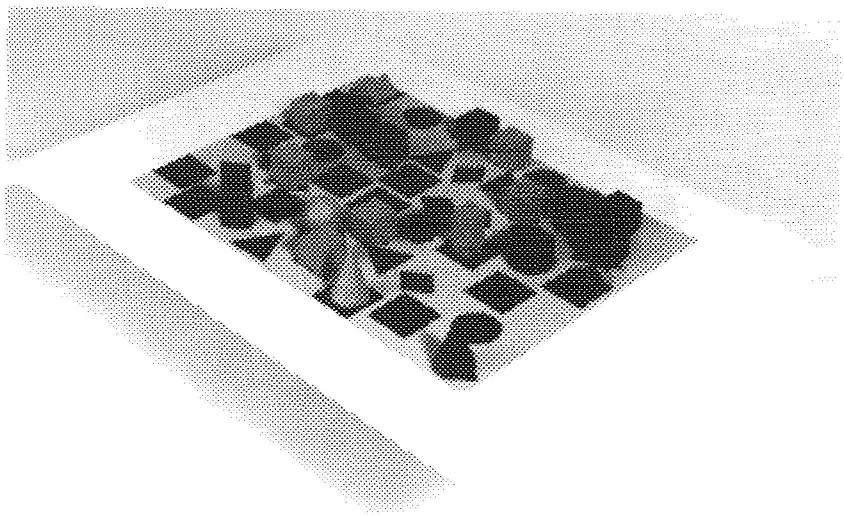

FIG. 2. Constellation of wooden pieces on the control board.

The *control board reconstruction task* was designed to provide a control task more comparable to the chess board reconstruction task than the digit-span control task used by Chi (1978). Figure 2 shows one of the two arrangements used for this task. There were 22 wooden pieces of two colors (yellow and blue) arranged on a board consisting of 48 different fields. The wooden pieces represented six different geometric figures (cylinder, cone, sphere, prism, cube, cuboid) with equal numbers of blue and yellow items. Thus the task structure was superficially similar to that of the chess board reconstruction task. Immediate recall and repeated recall were assessed as described above. Delayed recall was not measured for this task.

*Digit span task.* A standard digit span task, the German version of the WISC (Hamburg–Wechsler Intelligence Test; see Tewes, 1985) was used to assess subjects' memory span. Randomized lists of 3 to 10 digits were read at a rate of one digit per second. The first list had three digits; the number was increased successively until the subject was unable to reproduce the sequence of digits in the correct order on two consecutive trials. The subject's span was the largest number of digits correctly reproduced.

*Knight's Tour task.* In the original Knight's Tour task (Chase & Simon, 1973; Holding, 1985), subjects were given a chess board with four black pawns and a white knight positioned in the lower left corner. Subjects needed to move the knight in legal ways so that it landed on all squares except those with a pawn or controlled by a pawn. A slightly modified version of this task was used in our study mainly to simplify the instruction and make the task easier for the child and adult novices. That is, the knight had to be moved through not more than one rank, and was also allowed to land on squares controlled by a pawn.

Time to complete this task was chosen as the dependent variable because it has been shown to be a valid measure of chess knowledge (Chi, 1978; Radojcic, 1971). We did not use number of errors in our analyses because it was generally low even in the novices.

*Interview.* Children were asked about their age, grade, school level (e.g., high, middle, or low educational track), and grade average in language (German) and math during their last school year. Most children (about 80%) attended the gymnasium, that is, the high educational track. They also were asked whether they belonged to a chess club, how often they practiced and participated in competitions, and when they started playing chess. The procedure for the adults was identical. As all adult subjects were linked to the university system, only the grade average of their school leaving exam (Abitur) was recorded.

## RESULTS

### *Knight's Tour*

A 2(expertise) × 2(age) analysis of variance using total time as the dependent variable yielded a significant main effect of expertise, $F(1, 76) = 18.57$, $p < .01$), no effect of age, and no significant interaction. The mean completion times for the adult and child experts were 133.3 and 180.5 s, respectively. In comparison, the adult and child novices were considerably slower (224.5 and 249.8 s, respectively). Subsequent Student–Newman–Keuls tests revealed that the child and adult experts were significantly faster than the child and adult novices (all $p$'s $< .05$). These findings thus corroborate our classification of experts and novices described above.

### *Memory Performance*

*Reconstruction of meaningful chess positions.* Preliminary analyses revealed that performance on the two meaningful chess reconstruction tasks (i.e., immediate recall and recall across trials) did not differ within the expert and novice groups. Thus mean performance on these two tasks was used as the dependent variable in subsequent analyses.

Table 1 shows the mean number of items correctly reconstructed separately for the three recall conditions (i.e., immediate recall, repeated recall, and delayed recall) as a function of expertise and age. A 2(expertise) × 2(age) × 5(trials) repeated measurement analysis of variance on these data (i.e., on trials 1 to 5) revealed a significant effect of expertise, $F(1, 76) = 61.12$, $p < .01$ and of trials, $F(4, 304) = 611.05$, $p < .01$. There was no effect of age, but a significant expertise × trial

TABLE 1
MEAN NUMBER OF CHESS PIECES CORRECTLY RECONSTRUCTED IN MEANINGFUL POSITIONS ACROSS TRIALS, AS A FUNCTION OF EXPERTISE AND AGE

| | Group | | | |
|---|---|---|---|---|
| | Children | | Adults | |
| Condition | Experts | Novices | Experts | Novices |
| Immediate recall | 8.87 | 4.95 | 7.10 | 4.58 |
| | (2.91) | (1.68) | (2.51) | (1.43) |
| Trial 2 | 13.00 | 7.85 | 12.53 | 7.43 |
| | (4.14) | (2.15) | (4.55) | (2.13) |
| Trial 3 | 16.23 | 10.20 | 16.25 | 9.53 |
| | (3.74) | (2.71) | (3.93) | (2.95) |
| Trial 4 | 18.58 | 13.35 | 19.00 | 12.90 |
| | (3.32) | (3.58) | (3.22) | (4.03) |
| Trial 5 | 19.52 | 15.90 | 20.85 | 15.28 |
| | (3.02) | (3.80) | (1.91) | (4.71) |
| Delayed recall | 11.20 | 4.95 | 11.35 | 4.20 |
| | (6.32) | (4.36) | (8.00) | (4.53) |

*Note.* Standard deviations in parentheses.

interaction, $F(4, 304) = 9.63, p < .01$. Experts gained consistently more than novices during the first four trials, whereas novices gained more than experts from the fourth to the fifth trial due to the fact that experts operated close to ceiling on trials 4 and 5. A trend analysis of correctly reconstructed items resulted in a significant linear trend as a function of expertise, $F(1, 76) = 980.09, p < .01$, and a significant quadratic trend, $F(1, 76) = 37.63, p < .01$.[1]

An additional repeated measures analysis of variance on immediate and delayed recall revealed significant effects of expertise, $F(1, 76) = 32.54$, $p < .01$, and recall condition, $F(1, 76) = 8.96, p < .01$. These main effects were qualified by a significant expertise × recall condition interaction, $F(1, 76) = 7.50, p < .01$. Whereas experts' recall was significantly better on the delayed than on the immediate recall condition, the novices performed similarly on both occasions. There were no effects of age, and

[1] Once a subject attained a perfect score and the trials were ended, a score of 22 was entered on all subsequent trials for this subject. Counting the experts as generating perfect performance on subsequent trials probably biases the estimate of performance positively, since one perfect performance does not guarantee later perfect reconstructions. On the other hand, not counting them would bias the estimates of experts' performance negatively, since they probably would have continued to do well. Omitting those experts from analysis who reached ceiling before the last trial would have lead to a considerable loss of subjects in that group. Thus we decided to keep these subjects despite the risk of positive bias described above.

TABLE 2
MEAN NUMBER OF CHESS PIECES CORRECTLY RECONSTRUCTED IN THE RANDOM POSITION ACROSS TRIALS, AS A FUNCTION OF EXPERTISE AND AGE

| | Group | | | |
|---|---|---|---|---|
| | Children | | Adults | |
| Condition | Experts | Novices | Experts | Novices |
| Immediate recall | 5.00 | 3.25 | 3.60 | 2.70 |
| | (2.20) | (1.92) | (2.04) | (1.63) |
| Trial 2 | 8.50 | 5.56 | 6.40 | 5.00 |
| | (2.16) | (2.12) | (2.26) | (2.58) |
| Trial 3 | 11.85 | 8.40 | 9.95 | 7.80 |
| | (3.15) | (2.84) | (2.70) | (3.19) |
| Trial 4 | 14.75 | 10.40 | 13.20 | 9.30 |
| | (3.67) | (3.30) | (3.97) | (3.67) |
| Trial 5 | 17.35 | 12.80 | 15.80 | 10.40 |
| | (4.42) | (3.86) | (4.07) | (4.37) |

*Note.* Standard deviations in parentheses.

no significant interactions related to age. Expertise accounted for 32 and 24% of variance on immediate and delayed recall, respectively. These findings nicely confirm our expectation that performance on the meaningful chess reconstruction task (i.e., immediate recall, learning ability, and long-term retention) depends heavily on chess expertise, regardless of chronological age.

*Reconstruction of random chess positions.* Table 2 shows the mean number of items correctly reconstructed separately for the five trials as a function of expertise and age. A 2(expertise) × 2(age) × 5(trials) repeated measurement analysis of variance revealed significant effects of expertise, $F(1, 76) = 28.66$, $p < .01$, age, $F(1, 76) = 5.67$, $p < .05$, and trials, $F(4, 304) = 355.15$, $p < .01$. In addition, a significant expertise × trial interaction was obtained, $F(4, 304) = 10.94$, $p < .01$. Experts outperformed novices on all trials, children were better than adults, and all groups improved significantly over trials. However, experts gained more than novices across trials. A trend analysis of correctly reconstructed items resulted in a significant linear trend, as a function of expertise, $F(1, 76) = 17.07$, $p < .01$, and a nonsignificant quadratic trend, $p > .05$. Although the effects of expertise on performance in this task were less strong than those obtained for the reconstruction of meaningful chess positions, they were reliable for all dependent variables (i.e., immediate and repeated recall). Whereas expertise accounted for only 9% of the variance in immediate recall, it explained about 25% of the variance on the fifth trial.

TABLE 3
MEAN NUMBER OF ITEMS CORRECTLY RECONSTRUCTED ACROSS TRIALS ON THE CONTROL BOARD TASK, AS A FUNCTION OF EXPERTISE AND AGE

| | Group | | | |
|---|---|---|---|---|
| | Children | | Adults | |
| Condition | Experts | Novices | Experts | Novices |
| Immediate recall | 3.33 | 2.55 | 3.15 | 3.00 |
| | (0.88) | (0.99) | (1.13) | (1.29) |
| Trial 2 | 5.18 | 4.28 | 5.78 | 5.25 |
| | (1.88) | (1.32) | (1.63) | (1.51) |
| Trial 3 | 7.30 | 5.78 | 8.28 | 6.95 |
| | (3.16) | (1.59) | (2.11) | (2.54) |
| Trial 4 | 9.20 | 7.58 | 11.10 | 8.93 |
| | (2.96) | (2.21) | (2.77) | (2.83) |
| Trial 5 | 11.33 | 8.65 | 13.33 | 10.65 |
| | (4.12) | (2.97) | (3.41) | (3.66) |

*Note.* Standard deviations in parentheses.

*Control board reconstruction task.* Table 3 shows the mean number of items correctly reconstructed separately for the immmediate and repeated recall conditions, as a function of expertise and age. A 2(expertise × 2(age) × 5(trials) repeated measurement analysis of variance yielded significant effects of expertise, $F(1, 76) = 10.14$, $p < .01$, age, $F(1, 76) = 6.23$, $p < .05$, and trials, $F(4, 304) = 344.54$, $p < .01$. However, the main effects were qualified by a significant expertise × trials interaction, $F(4, 304) = 6.90$, $p < .01$, and a significant age × trials interaction, $F(4, 304) = 4.51$, $p < .05$. Simple-effects tests revealed that there were no significant effects of expertise and age in the immediate recall condition (all $p$'s > .05). However, experts improved more across trials than novices, and adults gained more across trials than children. Thus our expectation that expertise should not influence performance in this control task was borne out for immediate recall but not for learning.

*Digit span task.* A 2(expertise) × 2(age) analysis of variance of digit span performance yielded a significant main effect of age, $F(1, 76) = 31.84$, $p < .01$. There was no significant effect of expertise, and no significant interaction. Child experts' average memory span was 6.25 ($SD = 1.01$), as compared to a mean of 5.75 obtained for the child novices ($SD = .97$). Adult experts' average memory span was 7.65 ($SD = .93$) and slightly higher than that of adult novices ($M = 7.05$; $SD = 1.31$). Age accounted for 27% of the variance in the span measure, as compared to 3.75% attributable to chess expertise. This finding replicates Chi's (1978) findings.

*Multivariate prediction of memory performance.* In the analyses summarized above, expertise was treated as a dichotomous variable. One problem with such an approach is that a considerable amount of information in the (continuous) chess knowledge variable is ignored. To obtain a more accurate estimate of the impact of expertise on memory performance, multivariate regression analyses based on the total sample of 80 subjects were carried out for immediate recall in the two chess-related memory tasks (i.e., meaningful and random chess positions) and the board control task. In each analysis, age, educational attainment (i.e., grades), experience with chess (i.e., mean frequency of practice per year), digit span, and total time needed to complete the Knight's Tour task were regressed on the memory variable.

For immediate recall of meaningful positions, both the Knight's tour measure and experience with chess contributed significantly to the regression equation, accounting for about 40% of variance in performance. No other predictor variable made a significant impact. For recall of random chess positions, only experience with chess explained significant variance (14%). Finally, for performance on the board control task, only memory span made a significant impact, accounting for a small proportion of the variance (7%). Neither educational attainment nor chess knowledge affected performance on this control task. It should be noted that regression analyses conducted for recall on trials 2 to 5 yielded similar patterns of results. Details concerning these analyses are not provided because of space restrictions.

By and large, the findings from the regression analyses converge with the results from the analyses of variance reported above. Whereas expertise accounted for a large proportion of criterion variance in the meaningful chess reconstruction task, its impact was less pronounced in the random chess reconstruction task and negligible in the board control task.

*Strategic Behavior and Memory Performance*

The analyses described above reveal substantial expert–novice differences on the chess board reconstruction tasks, regardless of age group. The fact that these differences were particularly large for memory of *meaningful* chess positions illustrates the impact of domain-specific knowledge on memory performance related to the domain. However, as noted by Ornstein and Naus (1985), this finding does not explain *how* domain-specific knowledge exercises its impact.

We adopted two approaches to exploring this issue. The first is based on the assumption that experts are able to encode larger "chunks" than novices, that is, to encode several chess pieces as a single semantic unit of analysis, a meaningful pattern that is immediately recognized (cf. Chase & Simon, 1973). According to Chase and Simon, all chess pieces reconstructed within a 2-s interval belong to the same chunk. This criterion was also adopted for our study. As all of our sessions with the children were videotaped, it was possible to partition the recall data into chunks using a 2-s interresponse latency time (IRT) as the boundary. Because of technical problems, the video data of two children could not be analyzed. Given the enormous amount of effort and time required for the video analysis, only immediate recall for one of the meaningful chess positions was considered.

The second measure of "chunking" was the "collective reconstruction" measure developed by Bratko, Tancig, and Tancig (1986). This measure reflects the relative frequencies with which particular chess pieces are reconstructed, rather than interresponse latencies. The assumption is that if the same chess pieces are memorized together by most players, they seem to belong to the same chunk used by these players when memorizing the chess position.

*Interresponse latencies.* The first trial data of the repeated recall task were partitioned into chunks using a 2-s IRT as the boundary. From this analysis, we found that child experts retrieved an average of 6.1 chunks, whereas child novices retrieved an average of 4.6 chunks for the first trial, $F(1, 36) = 3.9$, $p > .10$. Although the experts tended to form more chunks than the novices, this difference was not reliable. Further, the average size of the chunks found for the experts (2.0) was not significantly larger than that found for the novices (1.8), $F(1, 36) = .9$, $p > .05$. Thus our findings do not replicate those by Chase and Simon (1973) and Chi (1978) who found that better players retrieve more and larger chunks in a single recall trial.

A closer inspection of our videotapes suggested that the 2-s interval was not ideal for identifying chunking processes in our subjects. To illustrate the problem with choosing the optimal boundary, data were reanalyzed using arbitrary IRTs that ranged from 880 to 3400 ms. As can be seen from Fig. 3, an IRT of 880 ms corresponds to 25% of observed frequencies of interresponse latencies, whereas an IRT of 3400 ms corresponds to 75% of observed frequencies of interresponse latencies. The classic 2-s IRT boundary corresponds exactly with the median of the distribution.

Significant differences in the "chunk" size of experts and novices were found for four of five IRTs (i.e., for all IRTs except the 2-s boundary[2]).

[2] This finding is not obvious from Fig. 3 which only gives the differences in means. For example, the mean chunk sizes for the experts and novices obtained for the 880 ms boundary were 1.37 and 1.20, respectively. Standard deviations were low in both cases (.26 and .20, respectively). For the 2-s IRT boundary, differences in mean chunk sizes were comparable for experts and novices (2.02 and 1.78, respectively). However, the standard deviations were considerably larger (.55 and 1.0, respectively). It is due to this difference in standard deviations that the mean differences in chunk sizes turned out to be significant for the 880 ms boundary but not for the 2-s threshold.

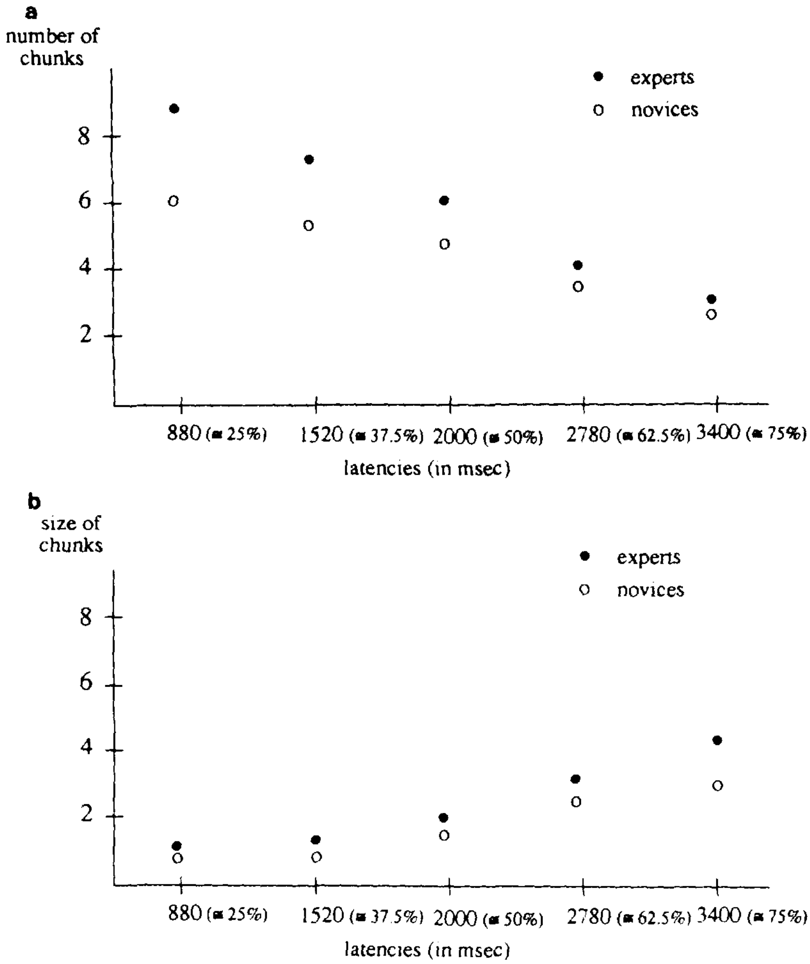


FIG. 3. Number and size of chunks obtained for various interresponse latencies, as a function of expertise.

On the other hand, the number of chunks did not differentiate as well between experts and novices: Only two of five comparisons (i.e., for 880 and 1520 ms) yielded significant results.

More important, the findings in Fig. 3 show a systematic, linear relationship between the IRT chosen, the size, and the number of chunks identified in the analyses. That is, the smaller the IRT unit, the smaller the size and the larger the number of chunks. The findings regarding chunk size suggest that differences among experts and novices increase with increasing IRT boundaries. This could imply that activating rich domain knowledge takes some time. On the other hand, our results con-

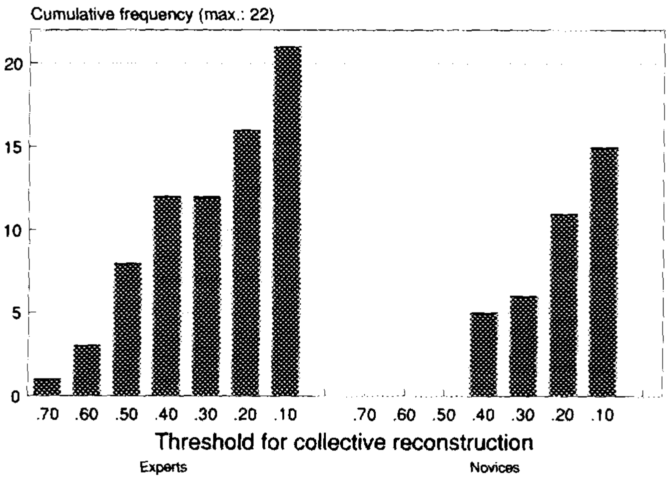


FIG. 4. Number of chess pieces successfully placed by 10 to 70% of subjects, separately for experts and novices.

cerning the number of chunks demonstrate an inverse relationship: Differences between experts and novices are most pronounced for low IRT boundaries and decrease with increasing time intervals. It appears, then, that speed of encoding and retrieval processes also make a difference between experts and novices. This finding corresponds well with the results obtained by Roth (1983).

These analyses suggest that adopting a 2-s IRT is as arbitrary as choosing one of the other intervals. Without a substantive reason for choosing a particular length, the standard choice of 2 s does not seem well justified.

*Collective reconstruction.* As noted above, the chunking measure developed by Bratko *et al.* (1986) seems better suited to assessing qualitative differences in experts' and novices' reconstruction strategies in that it takes into account the frequency with which particular chess pieces were used during the reconstruction trials. For example, the white king was correctly reconstructed by about 70% of the (child and adult) experts but less than 50% of the novices. Figure 4 shows the cumulative totals, that is, the number of pieces successfully placed by 10 to 70% of the subjects, separately for experts and novices. Bratko *et al.* (1986) concluded from their experiments that adjusting the threshold for each position to the value which maximizes the total number of pieces in all derived chunks lead to a threshold of about 0.4. This threshold was reported to be stable over all positions. When the threshold of collective reconstruction for our data was set to 0.4 (which means that a piece was collectively reconstructed if it was reconstructed by at least 40% of subjects), 12 pieces were found to be collectively reconstructed by the experts, as compared to only 5 pieces collectively reconstructed by the novices. This finding indicates that the majority of experts—but only a few novices—seem to follow a specific plan when reconstructing the meaningful chess position. More specifically, 10 of the 12 chess pieces collectively reconstructed by the experts during immediate recall built up the white and black King's wing positions, whereas the two remaining pieces belonged to the white Queen's wing. This pattern was completely reconstructed by all experts at the third trial. In comparison, the five white pieces collectively reconstructed by the novices during immediate recall did not make up a familiar position.

Although the collective reconstruction procedure suggested by Bratko *et al.* (1986) is only descriptive in nature, it gives valuable information about qualitative differences in the reconstruction strategies of experts and novices. The findings from this analysis were also confirmed by a close inspection of the videotapes which showed that most experts seemed to follow a similar plan when reconstructing the first chess pieces from memory, whereas the novices' initial reconstruction patterns were heterogeneous and unpredictable.

## DISCUSSION

One major goal of the present study was to replicate Chi's (1978) classic results. We were able to confirm several of Chi's findings. First of all, we replicated Chi's most impressive finding that child experts' immediate recall for meaningful chess positions was far superior and significantly better than that of adult novices. Although different chess positions were used in the two studies, differences in performance found between child experts and adults novices were comparable (9.3 pieces versus 5.9 pieces in Chi's study, as compared to 8.8 versus 4.6 in our study). Second, we replicated Chi's finding that the child experts' digit spans were lower than those of the adult novices (means of 6.1 digits versus 7.8 digits in Chi's study, as compared to 5.8 versus 7.6 in our study). Due to the considerably larger sample size in the present study, we were able to show that this difference is highly reliable. Taken together, the results of both studies nicely demonstrate that rich knowledge in a specific domain strongly affects memory performance on tasks dealing with that specific domain, thereby leading to a reversal of typical age trends.

Another important goal of our study was to extend the design of Chi's study in several aspects. First, subgroups of adult experts and child novices were included in our study in order to obtain unconfounded estimates of the impact of domain knowledge and chronological age on memory performance and behavior. Second, three variations of the same task were used to explore the impact of familiarity with the chess pieces and the chess board. Our basic assumption was that experts' superiority should be most evident in cases in which familiarity with both deep and superficial characteristics can influence results (i.e., the meaningful chess position).

The impact of knowledge on performance should be less pronounced but still significant if only one of the two familiarity components, that is, familiarity with the geometrical pattern of the board and the form and color of chess pieces, can be assumed to be effective (i.e., for random chess positions). Finally, performance differences between experts and novices should be eliminated when both familiarity components are no longer available (i.e., the control board).

The results supported such a view, at least as far as immediate recall is concerned. That is, expertise was the most important source of individual differences in the meaningful chess board reconstruction task, accounting for 32% of variance in immediate recall. In comparison, age explained only 3%. As expected, the impact of expertise on recall of the random chess position was still significant but less pronounced, accounting for only 9% of variance. Again, the contribution of age differences was not reliable (about 4% of the variance were explained by this variable). Also in accord with our predictions, neither expertise nor age contributed significantly to immediate recall on the control board task. Both the ANOVAs and the multiple regression analyses showed these patterns.

The pattern of results for long-term retention (i.e., delayed recall) of meaningful chess positions was similar to that obtained for immediate recall: Whereas expertise accounted for a significant proportion of the variance in delayed recall (about 24%), the impact of chronological age was negligible. Inspection of Table 1 shows that the novices' delayed recall corresponds to their immediate recall, whereas both expert groups recalled more items in delayed than in immediate recall. This finding supports Charness' (1976) observation that chess experts store information acquired during a short exposure to chess piece constellations in long-term memory, where it is quite resistant to interference.

The analysis of repeated recall yielded interesting insights concerning the roles of expertise and age on achievement gains across trials. On both the meaningful and random chess positions, a significant expertise × trials interaction indicated that experts not only performed better on the first trial (i.e., immediate recall) but also gained more than novices on subsequent trials. This finding demonstrates that expertise can have substantial effects on the quality and speed of the learning process. A comparison of recall on the first and last trials of the random chess position illustrates this phenomenon particularly well: Whereas expertise did not account for more than 9% of variance in immediate recall, it explained 25% on the last learning trial. The findings for the control board task paralleled those for the random positions in that experts gained more than novices across trials. However, the findings differed from those obtained for the random chess position in that age made an independent impact: The achievement gains of adults were greater than those of the children. Although we do not have a good explanation for the latter finding, there is evidence in the literature that adults learn faster than children across trials when the task materials are unfamiliar to the subjects (cf. Lindberg, 1980).

Given these reliable age differences in favor of adults for the control board task, it seems surprising that children outperformed adults on the random chess board task, and that no age differences among experts on expertise-related materials were found. In our view, these unexpected findings are due to the fact that we did not succeed in matching the level of expertise in children and adults. Our interview data revealed that the overall level of chess expertise was higher in the child novices than in the adult novices. On average, child novices had started playing chess at an earlier age than the adult novices and used to practice more than the adults. Similar differences were found between child and adult experts. Whereas most of the children were tournament players, only a few of the adult experts had continued to play chess regularly. On average, the child experts had started playing chess at an earlier age than the adults (8 years vs. 14 years, respectively). They also practiced a lot more than the adults ever did. Probably due to this difference in level of expertise, children tended to be better than adults on immediate recall of the meaningful chess positions, and clearly outperformed the adults in the random chess board task where the meaningfulness of the chess board pattern was reduced.

The difference in level of expertise may also explain the absence of age differences among experts on expertise-related materials. Although the adult experts' chess skills were clearly above average, those of the child experts were outstanding. It is probably true that children who learn to play chess almost perfectly within a short period of time possess more "native expertise," that is, more of the specific memory abilities critical to chess than adult players who built up their expertise rather slowly over a period of many years. Presumably the children's advantage regarding native expertise did compensate for their lack of experience with chess.

A last goal of our study was to analyze qualitative differences in experts' and novices' recall. In particular, we were interested in replicating Chi's finding that experts recall more chunks and more pieces per chunk than novices. As indicated above, we experienced problems with adopting Chi's boundary for the definition of chunks (i.e., a 2-s IRT). The data from our memory experiment showed systematic relationships between the interresponse latency chosen and the number and size of chunks identified. This makes it difficult to trust the outcomes of such analyses. Although the results showed a reliable relation between the size of the first chunk and expertise for most IRTs considered in our analyses, one should be cautious in interpreting these findings. Some of the problems were already discussed by Chi, who concluded that the technique of partitioning chunks by a 2-s time interval cannot capture the complete structure of a chunk, because chunks may actually be overlapping and related and may be of different structures and sizes for players of different skill.

We believe that the problem of identifying chunks is particularly serious in the reconstructive memory task used by Chi and in this study. Please note that the original procedure used by Chase and Simon (1973) to assess chunking was different in that subjects did not have to rely on memory: They were allowed to look back at the target board as needed. Using such a procedure and the number of glances required to reconstruct the position as the dependent variable, Horgan and Morgan (1990) were able to show that number of look-backs correlated significantly with the rating of their chess experts. However, although this procedure seems better suited to assessing the size and number of chunks, it does not give any information on the structure of contents.

In our view, the major advantage of the collective reconstruction index developed by Bratko *et al.* (1986) is that it seems suited to identify qualitative differences in the chunk structure of experts and novices. We were able to show that experts and novices behave differently when reconstructing the chess board. While most experts seem to start with the reconstruction of specific meaningful units, the novices seem to focus on aspects like color of pieces or specific positions on the board, regardless of age. Although more sophisticated techniques need to be developed to test the validity and generalizability of this technique, it seems to be a step in the right direction.

Taken together, our results suggest that chess experts outperform chess novices on a chess board reconstruction task because of their greater familiarity with (a) meaningful constellations of chess pieces and (b) the geometrical pattern of the board and the form and color of chess pieces. Experts' knowledge of these factors enables them to process information faster and in larger semantic units, which allows them to use different strategies and to remember and learn more than novices in this domain of expertise.

## REFERENCES


Anderson, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). New York: Freeman.

Bjorklund, D. F. (1990). *Children's strategies: Contemporary views of cognitive development.* Hillsdale, NJ: Erlbaum.

Bratko, J., Tancig, P., & Tancig, S. (1986). Detection of positional patterns in chess. In D. F. Beal (Ed.), *Advances in computer chess* (Vol. 4, pp. 113–126). Oxford: Pergamon.

Charness (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory,* **2,** 641–653.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology,* **4,** 55–81.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Chi, M. T. H. & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 91–142). Orlando: Academic Press.

de Groot, A. D. (1946). *Het denken van den schaker* Amsterdam: Noord Hollandsche.

de Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). The Hague: Mouton.

Ericsson, K. A., & Crutcher, R. J. (1990). The nature of exceptional performance. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 10, pp. 187–217). Hillsdale, NJ: Erlbaum.

Holding, D. H. (1985). *The psychology of chess skill.* Hillsdale, NJ: Erlbaum.

Holding, D. H., & Reynolds, R. I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory and Cognition,* **10,** 237–242.

Horgan, D. D., & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology,* **4,** 109–128.

Lindberg, M. (1980). The role of knowledge structures in the ontogeny of learning. *Journal of Experimental Child Psychology,* **30,** 401–410.

Lories, G. (1987). Recall of random and non random chess positions in strong and weak chess players. *Psychologica Belgica,* **27,** 153–159.

Ornstein, P. A., & Naus, M. J. (1985). Effects of the knowledge base on children's memory strategies. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 19, pp. 113–148). Orlando: Academic Press.

Reynolds, R. I. (1982). Search heuristics of chess players of different caliber. *American Journal of Psychology,* **95,** 383–392.

Radojcic, M. (1971). What is your chess IQ? *Chess Life and Review,* **26,** 709–710.

Roth, C. (1983). Factors affecting developmental changes in the speed of processing. *Journal of Experimental Child Psychology,* **35,** 509–528.

Schneider, W., & Hasselhorn, M. (in press). Situational context features and early memory development: Insights from replications of Istomina's experiment. In R. van der Veer, M. van IJzendoorn, & J. Valsiner (Eds.), *Reconstructing the mind. Replicability in research on human development.* New Jersey: Ablex.

Schneider, W., & Pressley, M. (1989). *Memory development between 2 and 20.* New York: Springer-Verlag.

Siegler, R. S. (1991). *Children's thinking* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Tewes, R. (1985). *HAWIK-R: Hamburg-Wechsler-Intelligenztest für Kinder* (3rd ed.). Bern, Switzerland: Huber-Verlag.

RECEIVED August 6, 1992; REVISED January 4, 1993