Journal of the International Neuropsychological Society (1997), **3**, 598–607.

Concept formation and problem-solving following closed head injury in children

HARVEY S. LEVIN,[1] JAMES SONG,[1] RANDALL S. SCHEIBEL,[1] JACK M. FLETCHER,[2] HARRIET HARWARD,[3] MATT LILLY,[4] AND FELICIA GOLDSTEIN[5]

[1]Cognitive Neuroscience Laboratory, Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, TX
[2]Department of Pediatrics, University of Texas Medical School, Houston, TX
[3]Callier Center for Communication Disorders, University of Texas at Dallas, Dallas, TX
[4]Private consultant, Baltimore, MD
[5]Department of Neurology, Emory University, Atlanta, GA

Abstract

To further investigate the usefulness of 3 purported measures of executive function (EF) in head injured children, we administered the Twenty Questions Test (TQT), Tower of London (TOL), and the Wisconsin Card Sorting Test (WCST) to 151 children who had sustained a closed head injury (CHI) of varying severity about 3 years earlier. In addition, we tested 89 normal controls. Fifty-seven of the patients were included in a longitudinal study that compared performance at 3 months and 36 months. All of the head injured children underwent magnetic resonance imaging for investigational purposes. Severity of CHI, as defined by the lowest Glasgow Coma Scale (GCS) score, affected performance on all 3 EF measures. Focal lesion volume incremented prediction of performance on TOL and WCST, but not TQT. Moderate intercorrelations of the test variables were obtained. Although all three EF measures depicted changes in performance over 3 years, a ceiling effect detracted from the sensitivity of the TOL to the impact of CHI on development. Implications of the findings for clinical applications are discussed.
(*JINS*, 1997, *3*, 598–607.)

Keywords: Cognition, Head injury, Children, Magnetic resonance imaging

INTRODUCTION

Cognitive development in children is characterized by changes in their strategy of information seeking. As reviewed by Mosher and Hornsby (1966), the information obtained in the Twenty Questions Test (TQT) is determined by the type of questions asked and the individual's use of sequential answers to successively constrain the alternatives. In this procedure, the child views an array of pictures that belong to different categories such as animals and toys, and must determine which item the examiner has selected. Constraint-type questions, which eliminate two or more alternatives, depend on the child's ability to identify a superordinate category (e.g., animals) which relates to exemplars included in the display of items (e.g., cat, dog). Questions that impose constraints reflect a conceptual scheme and a hierarchical organization of information. In contrast, hypothesis or identity-seeking questions (e.g., "Is it the dog?") eliminate only a single alternative. Mosher and Hornsby (1966) also identified a pseudoconstraint question (e.g., "Does it bark?") which they described as eliminating only a single alternative despite appearing to be more conceptual than the hypothesis type of question. Goldstein and Levin (1991) found that long-term adult survivors of severe CHI tended to ask primarily hypothesis questions, a finding that the investigators interpreted as evidence for a conceptual impairment.

Developmental studies utilizing the TQT have reported that the proportion of constraint-type questions increases with age from 6 to 16 years (Levin et al., 1991; Mosher & Hornsby, 1966), and is paralleled by a decline in the proportion of hypothesis questions. Flexibility in conceptualizing the shared properties of items to formulate questions and utilizing feedback from the examiner to ask additional questions can be construed as an executive function (EF) which depends on the maturation and integrity of the pre-

Reprint requests to: Harvey S. Levin, PM&R Research Office, Baylor College of Medicine, 1333 Moursund Avenue, Rm. A205, Houston, TX 77030.

https://doi.org/10.1017/S1355617797005985 Published online by Cambridge University Press

frontal region. In view of the deleterious effects of CHI in children on cognitive skills that are widely interpreted as EFs (Levin et al., 1993; 1994), we postulated that severe injury would impair the development of a sophisticated question-asking strategy. Consequently, the primary goals of this study were to evaluate the effects of CHI severity on TQT performance and investigate the relationship of this cognitive measure to more established, purported tests of EF. To accomplish these goals, we administered all three cognitive tests to a cross-sectional cohort of head injured patients that included a subgroup whom we also studied longitudinally. We hypothesized that the TQT would have at least a moderate relationship with both the TOL and WCST, because all three measures assess flexibility in problem-solving, including utilization of feedback to change response strategy.

We studied the convergence of the TQT with other measures of EF, including the Tower of London (TOL) of Shallice (1982) and a computerized version of the Wisconsin Card Sorting Test (Grant & Berg, 1948; Heaton et al., 1993). The relationship of these widely accepted measures of EF to the TQT is largely unknown. We selected the TOL based on our recent principal components analysis (Levin et al., 1996) of EF measures which were given to 81 CHI patients and 102 normal children. A five-factor solution was obtained, including a factor on which a Twenty Questions variable and a TOL variable loaded. We interpreted this finding as the capacity to conceptualize a "schema" to guide problem solving. Although Levin et al. (1996) suggested that measures of the efficiency of asking questions and solving the TOL problems were tapping a common dimension, they also acknowledged the need for replication. Interpretation of the TOL as a measure of EF development received support from a cross-sectional study by P. Anderson et al. (1996) who found that performance improved from age 7 to 13 years, and was moderately correlated with Controlled Oral Word Association. However, a recent report by Cockburn (1995) indicated the relative insensitivity of the TOL to the long term cognitive sequelae of CHI in adults who exhibited deficits on word fluency and a modified Wisconsin Card Sorting Test (WCST). In the present study we also included the WCST because it is generally regarded as a "gold standard" of problem solving skills that tap working memory and is sensitive to prefrontal lesions despite false negative and false positive errors (S. Anderson et al., 1991). Based on neuropathologic (Adams et al., 1980) and neuroimaging (Mendelsohn et al., 1992) evidence for the selective vulnerability of the prefrontal region to focal lesions associated with CHI, our investigation of the TQT encompassed an analysis of focal brain lesions in head injured children. For comparison, we also studied the contribution of focal brain lesions to cognitive performance on the TOL and the WCST.

METHODS

Study Population

The sample included patients who had been hospitalized for CHI of varying severity which was defined by the lowest postresuscitation score on the Glasgow Coma Scale (GCS) of Teasdale and Jennett (1974). Selection criteria included (1) age 5 to 18 years at the time of testing; (2) nonpenetrating head trauma due to sudden acceleration or deceleration of the freely moving head, or being struck with a blunt object; and (3) no preinjury history of diagnosed neurologic or psychiatric disorder. Exclusion criteria included (1) injury due to child abuse; (2) a history of substance abuse, mental retardation, or learning disability; and (3) previous head injury resulting in hospitalization. All patients were recruited from consecutive admissions to neurosurgery services.

Table 1A summarizes the demographic and clinical features of the CHI groups in the cross-sectional study who had been admitted to Children's or Parkland Hospitals in Dallas, Texas, Hermann Hospital in Houston, Texas, the John Sealy Hospital in Galveston, Texas or the Johns Hopkins Hospital in Baltimore, Maryland. To assess the effects of CHI severity on question asking strategy, we compared the performance of 88 children who sustained *severe CHI* (i.e.,

Table 1A. Demographic and clinical features of the head injured and control groups in cross-sectional study

	Controls (N = 89)		Mild CHI (N = 63)		Severe CHI (N = 88)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age at study (years)	11.65	3.11	12.31	3.38	12.30	3.58
Age at injury (years)	-	-	7.94	3.86	7.27	3.69
Injury–study interval (months)	-	-	52.38	24.13	60.29	29.31
Parental education (years)	14.11	2.63	14.44	2.72	13.80	1.95
GCS score	-	-	14.48	0.69	5.57	1.69
Sex						
% Boys	61.80%		69.84%		59.09%	
% Girls	38.20%		30.16%		40.91%	

https://doi.org/10.1017/S1355617797005985 Published online by Cambridge University Press

GCS score 3–8) to 63 pediatric patients who were hospitalized for a *mild CHI*, which we defined as a GCS score from 13 to 15, duration of unconsciousness less than 30 min, no brain lesion on computed tomography (CT) within 24 hr of injury, and no focal brain lesion on magnetic resonance imaging (MRI) performed as part of this study. Eighty-nine neurologically intact children residing in the Dallas–Fort Worth metropolitan area, who were recruited through advertising on community bulletin boards and contacting local organizations, participated in the cross-sectional study to provide comparison data. The normal control participants were selected to approximate the demographic features of the CHI patients. As summarized in Table 1A, there were no group differences in age at the time of the study, sex, or parental education. Separate comparison of the mild and severe CHI groups disclosed no differences in the age at injury or at the time of this study. The interval from the date of injury to the time of study tended to be longer in the severe CHI group as compared to mildly injured patients (Table 1A). Implicit in the selection criteria for the two CHI groups, there was no overlap in the GCS scores. Severe CHI was most frequently produced by motor vehicle crashes (38 patients, or 43% of group) and pedestrian motor vehicle injury (31 patients, or 35% of group) as compared to the mild CHI patients [13 (21%) involved in motor vehicle crashes; 6 (10%) struck by motor vehicles]. In contrast, falls were more common in the mild CHI group ($n = 16$; 25%) relative to severely injured children ($n = 3$; 3% of group).

Table 1B shows the corresponding demographic and clinical features for the subset of head injured children who were studied longitudinally at both 3 and 36 months after sustaining a CHI. Comparison of the mild versus severe CHI groups who completed both the 3-month and 36-month assessments in the longitudinal study revealed no differences in demographic features or time from injury to the second assessment. The longer interval from injury to the initial test in the severe CHI group (Table 1B) reflects their prolonged disturbance of consciousness, injury complications, and extracerebral injuries, which necessitated deferral of the first assessment to 6 or 12 months postinjury in selected cases.

Table 1B. Demographic and clinical features of the head injured groups in longitudinal study

	Mild CHI ($N = 27$)		Severe CHI ($N = 30$)	
Variable	*M*	*SD*	*M*	*SD*
Age at first evaluation	9.97	2.85	10.03	3.19
Age at second evaluation	12.76	2.87	12.74	3.24
Age at injury	9.68	2.85	9.64	3.26
Injury–first evaluation interval (months)	3.44	0.64	4.72	2.73
Injury–second evaluation interval (months)	36.99	1.67	37.25	1.91
Parental education (years)	13.89	2.81	13.43	1.87
GCS score	14.33	0.62	6.07	1.72
Sex				
% Boys	62.96%		53.33%	
% Girls	37.04%		46.67%	

Materials and Procedure

Magnetic resonance imaging (MRI)

Since the inception of the project in 1990, the MRI technology has evolved using various pulse sequences, thinner slices, and higher field magnets. However, the protocol has consistently included T1-weighted sagittal images, T1-weighted coronal images, and T2-weighted coronal images. Beginning in August 1991, patients ($N = 86$) were imaged in Dallas with a 1.5T Picker magnet (Picker International, Highland Heights, OH) to obtain 5-mm 3DFT T1-weighted sagittal and coronal images. In addition, 5-mm T2-weighted coronal images were done with no gap. All of the MRI scans were reviewed by a neuroradiologist independent of the cognitive data. The findings were entered on a coding form that specified the location of each focal area of abnormal intensity and atrophy. Intracranial lesions were measured with a Jandel planimeter (Jandel Scientific; Rafael, CA) connected to a microcomputer (IBM; Armonk, NY). The area of each lesion was measured on successive slices, and summed to obtain a total volume. All brain lesions were traced on templates developed for MRI coronal slices (Damasio, 1991).

Twenty Questions Test

The display consisted of 42 colored line drawings of common objects (e.g., bee, carrot, airplane) used by Mosher and Hornsby (1966) that were placed on a 46 × 61 cm board. The figures were arranged on the display such that their shared conceptual features were noncontiguous. The child was first asked to name each picture and was corrected if an inappropriate label was supplied. Next, the following instructions adapted from Denney and Denney (1973) were given:

> We are going to play a question-asking game. I am thinking of one of these pictures, and you have to guess which one it is. The way to guess is by asking questions that I can answer “yes” or “no”–any question at all as long as I can answer it “yes” or “no.” So go ahead and ask a question and try to find out which one it is in the fewest number of questions. Remember to try to guess the picture in the fewest number of questions you can.

The goal of guessing the item in the fewest number of questions was emphasized. In view of the lack of interactions between number of trials and groups (head injured or controls) in a previous study of adults (Goldstein & Levin, 1991), the present study was limited to a single trial. The child was allowed up to 30 attempts to guess the picture. Different items were selected for the 3- and 36-month examinations.

Scoring

Questions were scored as hypothesis (asking about a specific item: “Is it the shoe?”); pseudoconstraint (asking about a specific item in a nondirect manner: “Does it have shoelaces?”); constraint (asking about a category that referred to two or more pictures and thus narrowed down the alternatives: “Is it an animal?”); and repeat or redundant questions. Each type of question was expressed as a percentage of total questions. The total number of questions asked was also analyzed.

Tower of London

A modified version of the TOL (Shallice, 1982) was administered, which consisted of three colored beads (one blue, one red, one green) presented on three pegs of varying length. The test materials included two TOL units (the test unit and model), each consisting of a 7.5 × 15 cm wooden base and three metal rods of varying length (3.5, 5, and 7.5 cm). A series of problems involved planning successive moves of beads from a standard initial position to one of 12 goal arrangements, displayed on the adjacent model. This version of the TOL consisted of 12 problems, including four problems at each of three levels of increasing complexity (*low*, *medium*, *high*), defined by the minimum number of moves needed to match the model (2–3, 4, or 5 moves). The simpler problems could be solved by directly transferring beads from an initial to a goal position, whereas more complex problems involved planning the correct sequence of moves.

The examiner explained to the child that the goal was to rearrange the beads to match the model using the minimum number of moves. Although the latency prior to initiating the first move on Trial 1 of each problem (initial planning time) was recorded, the instructions emphasized accuracy rather than speed of performance. A maximum of three trials was given to solve each problem. The number of broken rules incurred (e.g., attempting to transfer more than one bead at a time) was recorded for each child. The examiner reminded the child of the instructions after each broken rule. The TOL measures included the percentage of problems solved on Trial 1, the percentage of problems solved within the three trials given for each problem, and the number of broken rules.

Wisconsin Card Sorting Test

The WCST (Grant & Berg, 1948) was administered and scored using a portable computer with a 10 × 15 cm screen and standard instructions (Heaton et al., 1993). Consistent with our previous studies employing the WCST (Levin et al., 1993, 1996), we analyzed the percent conceptual responses, number of categories attained, and the percent perseverative errors.

Vineland Adaptive Behavior Scale

To assess the relationship of the cognitive measures to everyday functioning, we included the Vineland Adaptive Behavior Scale (Sparrow et al., 1984) composite (VABSC) percentile score in the analysis. To provide consistency with previous studies based on this ongoing project (Fletcher et al., 1990, 1996), we analyzed the percentile score rather than the standard score on the VABSC.

Statistical Analysis

Multivariate analysis of covariance (MANCOVA) was used to test group and age at test effects on the performance measures (age at injury was analyzed in the longitudinal study, which was confined to the CHI groups because controls were tested on a single occasion). Age was used as a covariate in all of the analyses. To determine whether the volume of focal brain lesion was related to cognitive function, we performed multiple regression analyses to explore whether this variable incremented the prediction of cognitive performance after first entering the severity of injury (indexed by the GCS score), the child’s age at injury, and the interaction of CHI severity with age. Consequently, a multiple regression was performed for each region of interest (i.e., left frontal, left extrafrontal, right frontal, right extrafrontal) to assess the incremental contribution of lesion volume to the GCS score, age at injury, and the interaction of these variables. Preliminary analysis revealed that some of the variables were mildly skewed and not normally distributed. However, the statistics are based on the F test, which is relatively robust to moderate departures from normality, particularly with respect to mild skewness. Since the assumption of equivalent slopes was not rejected in a preliminary analysis of the TQT and WCST, the interaction of group with age is fitted only for the TOL. In view of the large number of multiple regressions, we view these analyses as exploratory, and advise caution in interpreting the results due to the increased risk of a Type I error. The results of the multiple regressions follow the analysis of severity of CHI and age.

RESULTS

Cross-sectional Study

Twenty Questions Test

Table 2 indicates the mean number of questions to solution and the proportions of each type of question for the CHI and control groups. The pattern of findings in Table 2 reflects a tendency for the severe CHI group to ask a higher percentage of hypothesis questions and a lower percentage of constraint questions relative to the mild CHI and control groups. Consistent with the view that constraint and hypothesis types of questions represent different levels of conceptualization, we confined the multivariate model to these variables. A MANCOVA was performed to compare the proportions of constraint and hypothesis types of questions asked by the head injured and control groups, disclosing a significant overall effect of CHI severity, which was confirmed on the univariate analyses for the percentage of hypothesis and constraint questions (Table 2). Total num-

Table 2. Summary of multivariate analysis of covariance and analysis of covariance for testing effects of severity of injury and age at test on Twenty Questions Test (A), Wisconsin Card Sorting (B), and Tower of London (C) in cross-sectional study. Means are also given (D)

	Group (G)		Age (A)		G × A		Level		G × L		A × L	
Measure	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
A. 20 Questions Test												
MANCOVA												
% constraint, % hypothesis	4.09	.003	46.11	.0001								
ANCOVA												
% constraint	8.08	.0004	85.94	.0001								
% hypothesis	3.35	.04	76.49	.0001								
Total no. questions*	4.11	.02	12.12	.001								
B. Wisconsin Card Sorting Test												
MANCOVA												
% conceptual, no. of categories	2.89	.02	25.81	.0001								
ANCOVA												
% conceptual	5.87	.003	51.06	.0001								
No. of categories	3.96	.02	39.01	.0001								
% perseverative errors*	9.63	.0001	38.10	.0001								
C. Tower of London												
MANCOVA												
% solved on Trial 1, % solved in 3 trials, no. of broken rules	5.46	.0001	17.86	.0001	4.96	.0001	10.56	.0001			4.25	.0001
ANCOVA												
% solved on Trial 1	3.87	.02	63.10	.0001			39.95	.0001			6.30	.002
% solved in 3 trials	12.52	.0001	29.46	.0001	7.88	.0004	12.63	.0001	2.29	.06	6.27	.002
No. of broken rules	9.45	.0001	28.70	.0001	6.15	.002	1.73	.18				
IPT*	5.69	.004	.05	.82	9.85	.0001	1.82	.16			9.64	.0001

D. Summary of group means in cross-sectional study after adjusting for age at test.+

	20 Questions Test			WCST			Tower of London Test			
Group	% constr	% hypoth	Total quest	% concept	Total cat	% persev	% Trial 1	% 3 trials	Rules	IPT
Control	51.29[a]	35.49[a]	10.73[a]	64.64[a]	4.96[a]	12.45[a]	62.59[a]	96.43[a]	0.05[a]	4.22[ab]
Mild	48.74[b]	37.23	11.70	59.40	4.61	14.70[b]	66.98[ab]	98.07[b]	0.09[b]	5.48[a]
Severe	38.08[ab]	45.53[a]	12.90[a]	55.24[a]	4.32[a]	17.56[ab]	61.38[b]	93.33[ab]	0.17[ab]	5.14[b]

Note. Within columns, common superscript letters denote significant ($p \leq .05$) pairwise contrasts.
F tests in MANCOVA are based on Wilks' Lambda.
*not included in MANCOVA.
+WCST = Wisconsin Card Sorting Test; IPT = initial planning time; constr = constraint; concept = conceptual; hypoth = hypothesis; quest = questions.

ber of questions asked was analyzed by using a univariate model and was also found to have a significant effect for group. Pairwise contrasts of all groups indicated that the severe CHI group asked a lower percentage of constraint questions than the mildly injured children and normal control group (Table 2). The severe CHI group also asked more total questions and higher percent of hypothesis questions than controls (Table 2). In the multivariate and univariate models of TQT, age at testing had a significant effect on the percentages of constraint and hypothesis questions (Table 2). None of the Age × CHI Severity interactions reached significance.

Multiple regression model including lesion volume to predict twenty questions scores. Using focal brain lesion volume to explore the predictive usefulness of models which incorporated the severity of CHI indexed by the GCS score, age at injury, the interaction of injury severity with age, and the lesion volume on MRI, four sets of multiple regressions (i.e., left frontal, left extrafrontal, right frontal, right extrafrontal) were run for the percentage of constraint and hypothesis questions and the total number of questions asked. Although the models were predictive of the percentage constraint (23–45% of variance) and the percentage hypothesis (28–54% of variance) questions, lesion volume did not significantly increment the prediction of performance beyond the level obtained using the GCS score, age, and the interaction of these two variables.

Wisconsin Card Sorting Test

Severity of injury and age. As reflected in Table 2, the normal control children tended to perform at a higher level than both head injured groups. Pairwise comparisons indicated that the severe CHI patients had the lowest percent of conceptual responses, and attained the fewest categories relative to controls, and had the highest percent of perseverative errors compared to both the controls and the mild CHI groups. MANCOVA and ANCOVA models of the WCST variables revealed significant effects for group and age (Table 2), but there was no interaction.

Multiple regression model including lesion volume to predict Wisconsin Card Sort. Multiple regression analyses disclosed that the model including GCS score, age, the interaction of GCS score with age, and lesion volume resulted in significant prediction of card sorting performance on the percent conceptual level responses (10–37% of variance), number of categories (1–27% of variance), and percent perseverative errors (9–27% of variance). Incremental prediction of percent conceptual level responses was provided by lesion volume in the right frontal region ($p = .006$) and the left extrafrontal region ($p = .05$). Incremental prediction of both the total number of categories attained and the percent perseverative errors was provided by lesion volume in the left ($p = .03$) and right frontal ($p = .04$) regions. Incremental prediction of the percent perseverative errors was provided by the addition of left frontal ($p = .01$) and right frontal ($p = .01$) lesion volumes.

To summarize, the WCST was sensitive to both severity of injury and the age at study. In addition, the volume of lesions in the left and right prefrontal lobes and the left extrafrontal region incremented the prediction of card sorting performance beyond the level attained by the GCS score and age.

Tower of London

A MANCOVA was performed on the TOL measures (i.e., percentage of problems solved correctly on Trial 1, within three trials, and the number of broken rules) using the level of complexity as a within-subject variable (Table 2). Initial planning time was analyzed separately, because the TOL is an untimed task, and we did not view this variable as closely linked to the measures of problem solving. The MANCOVA indicated that three main factors, including injury group, age, and level of complexity were significant, a finding which was confirmed in the ANCOVA models (Table 2), with the exception of the number of broken rules. Pairwise comparisons indicated that the severely injured children solved a lower percentage of problems within three trials and had a greater number of broken rules than the mild CHI and control groups (Table 2). The mild CHI group solved a higher percentage of problems on Trial 1 than either the severe CHI or control group (Table 2).

The significant interaction of CHI group with age reflects the fact that young children who sustained a severe head injury had a disproportionate impairment of solving TOL problems relative to older children who had been in coma (Table 2). The univariate analyses confirmed the Age × CHI Severity interaction on the percentage of problems solved within three trials and the number of broken rules. In particular, the increased tendency to break rules while performing the TOL was particularly salient in young children, as reflected by the Age × Severity interaction. The other significant 2-way interaction in MANCOVA was age with level of complexity, which reflected that level of complexity that had a disproportionate effect on the performance by young children relative to older children. Young children had more difficulty in solving the high complexity problems than older children. The univariate analyses confirmed the Age × Level of Complexity on the percentage of problems solved on Trial 1 and the percentage of problems solved within three trials.

In the univariate model for initial planning time, both the Age × CHI Severity and Age × Level of Complexity interactions were still significant. Both interactions have the same interpretation as before. ANCOVA also disclosed that the effect of group on the initial planning time was significant. Table 2 reflects a tendency for head injured children to have longer latencies before initiating their first move as compared to controls.

Multiple regression model including lesion volume to predict Tower of London performance. Multiple regressions for each region of brain lesion disclosed that the overall models consisting of the GCS score, age, the GCS × Age interaction, and lesion volume were predictive of performance on the TOL. The percentage of problems solved on Trial 1 and the IPT were the TOL measures most consistently predicted by the overall models. Incremental prediction by brain lesion volume was confined to lesions situated in the left hemisphere, including the contribution of frontal lesions to the percent problems solved on Trial 1 ($p = .02$) and initial planning time ($p = .003$). The volume of left extrafrontal lesions contributed to the prediction of the number of broken rules ($p = .0005$).

In summary, the TOL was sensitive to the severity of CHI and the volume of left hemisphere lesion. Both sustaining CHI at an early age and increasing the complexity of the problems accentuated the effects of CHI severity on TOL performance. Although the minimum number of moves required for solution was related to the percentage of problems solved and the number of broken rules, complexity interacted with age but not with severity of injury. Latency to initiate the first move was increased in head-injured patients as compared to the normal controls.

Intercorrelations of cognitive variables

The Spearman intercorrelations were significant, and generally ranged between .36 and .46, corresponding to approximately 13 to 20% of shared variance. The highest correlation obtained was between the TQT (percent constraint questions) and WCST (percent conceptual level responses, $r = .46$, $p < .0001$). The percent constraint questions were also correlated with TOL variables (e.g., percent solved within three trials, $r = .39$, $p < .0001$).

Relationship between cognitive variables and adaptive behavior

To examine the prediction of the VABSC from the executive function measures, we first performed a multiple regression analysis on all of the variables listed in Table 2. This overall regression equation accounted for 21% of the variance in the VABSC percentile score. Entering subsets of the cognitive measures in a series of multiple regressions indicated that the most efficient prediction was obtained by using a model that consisted of the percent of hypothesis questions on the TQT and the percent of conceptual responses on the WCST. This model produced an R-square of .1874, accounting for nearly 19% of the variance in VABSC scores.

Longitudinal Study

Twenty Questions

Table 3 depicts the mean percent of constraint and hypothesis questions according to the occasion. Overall, the MANCOVA revealed significant effects of CHI severity, age, and occasion (Table 3). Although the effect of age was significant for the percent constraint and percent hypothesis questions, there were no interactions.

Wisconsin Card Sorting Test

Table 3 shows the percentage of conceptual level responses, number of categories, and percent perseverative errors for the initial and follow-up examinations. The results of MANCOVA and ANCOVA (Table 3) indicated that the effect of occasion was significant for all three WCST measures reflecting an overall improvement in performance. Table 3 also shows that the effect of age at injury was significant for all three WCST variables. The interaction of CHI severity with occasion was significant in the MANCOVA model and in the ANCOVA for the percentage of conceptual responses and number of categories. However, there was no significant effect of injury group in MANCOVA, nor did severity of injury interact with age.

Tower of London

The multivariate analysis confirmed the cross-sectional results in showing significant overall effects of CHI severity, level of complexity, and the interaction of Age × Level and Age × Severity on TOL performance. Occasion, and its interactions with CHI severity and age, were also significant.

Table 3 presents the mean scores for each occasion, and indicates that the interaction of Occasion × Age and Occasion × Severity Group were significant for percent problems solved in three trials, the number of broken rules, and the initial planning time. This reflects that the increased percent of problems solved in three trials and the reduction in broken rules and initial planning time over 3 years was greater in young children relative to older patients, and in more severely injured patients as compared to the mild CHI group. Severity of head injury had a more pronounced effect on the TOL performance of young children relative to older children, reflecting the Group × Age interaction for the number of broken rules and initial planning time (Table 3). Severe CHI patients disproportionately improved their performance over time relative to the children who sustained mild head injury, as supported by the significant interaction of Occasion × Severity Group in the univariate test for the number of broken rules. The number of broken rules declined at follow-up, as indicated by the significant effect of occasion in Table 3.

DISCUSSION

In view of the vulnerability of the prefrontal region to focal lesions associated with severe CHI (Mendelsohn et al., 1992) and the implications for development of EF in children, the usefulness of putative EF measures for outcome research and clinical application assumes importance. Consequently, we investigated the relationship of three EF measures to the severity of CHI, age at injury, and volume of focal brain lesions.

The TQT assesses the child's ability to formulate questions that reflect the semantic features common to two or

Table 3. Summary of multivariate analysis of covariance and analysis of covariance for testing effects of severity of injury, age at test, and occasion on Twenty Questions Test (A), Wisconsin Card Sorting Test (B), and Tower of London (C) in longitudinal study. Means are given (D)

	Group (G)		Age (A)		G × A		Occasion		G × O		Level		A × O		A × L	
Measure	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
A. 20 Questions Test																
MANCOVA																
% constraint, % hypothesis	5.43	.006	19.38	.0001			11.14	.0001								
ANCOVA																
% constraint	10.85	.001	29.71	.0001			22.47	.0001								
% hypothesis	3.31	.07	33.82	.0001			8.54	.004								
Total no. questions*	3.94	.05	1.56	.21			7.50	.01								
B. Wisconsin Card Sorting Test																
MANCOVA																
% conceptual, no. of categories	2.77	.07	13.76	.0001			8.38	.0004	3.34	.04						
ANCOVA																
% conceptual	5.32	.02	27.66	.0001			14.81	.0002	6.32	.01						
No. of categories	4.60	.03	19.27	.0001			15.53	.0001	5.64	.02						
% perseverative errors*	3.24	.07	24.6	.0001			10.37	.002	3.73	.06						
C. Tower of London																
MANCOVA																
% solved on Trial 1, % solved in 3 trials, no. of broken rules	5.29	.001	1.91	.13	3.25	.02	7.44	.0001	5.53	.001	16.93	.0001	4.49	.004	4.48	.004
ANCOVA																
% solved on Trial 1	7.99	.005	17.89	.0001			38.43	.0001			21.01	.0001			2.09	.13
% solved in 3 trials	8.26	.004	19.47	.0001	3.32	.07	16.58	.0001	6.63	.01	8.84	.0002	7.56	.01	4.89	.01
No. of broken rules	13.08	.0003	16.93	.0001	6.08	.01	20.45	.0001	14.94	.0001	.08	.92	11.63	.001		
IPT*	4.80	.03	0.91	.34	4.84	.03	9.26	.003	6.21	.01	0.62	.54	10.67	.001	2.16	.12

D. Summary of means in longitudinal study

	20 Questions Test			WCST			TOL			
Occasion	% constr	% hypoth	Total quest	% concept	Total cat	% persev	% Trial 1	% 3 trials	Rules	IPT
3 month	33.09[a]	44.86[a]	13.51[a]	50.32[a]	3.47[a]	19.66[a]	55.15[a]	90.05[a]	0.34[a]	5.39[a]
36 month	53.97[a]	29.67[a]	10.63[a]	65.48[a]	5.03[a]	12.70[a]	70.06[a]	98.02[a]	6.05[a]	5.49[a]

Note. Within columns, common superscript letters denote significant ($p \leq .05$) pairwise contrasts.
F tests in MANCOVA are based on Wilks' Lambda.
*not incuded in MANCOVA.
+WCST = Wisconsin Card Sorting Test; IPT = initial planning time; constr = constraint; concept = conceptual; hypoth = hypothesis; quest = questions.

more items illustrated on a display. The developmental shift from asking primarily hypothesis-type questions to formulating questions that impose constraints and eliminate several alternatives (Denney & Denney, 1973) presumably reflects prefrontal maturation. The capacity to utilize feedback from the examiner to reformulate constraint questions permits efficient elimination of items. Thus, inflexible problem-solving, perseveration, and difficulty in conceptualizing common features would impair question-asking strategy. In the present study we found that TQT performance was sensitive to CHI severity and age at injury, especially with regard to the proportion of constraint questions asked by the children. However, other variables in the TQT were not consistently sensitive to CHI severity, and performance was minimally related to the volume of prefrontal or extrafrontal brain lesions. The TQT was sensitive to the effect of occasion, indicating improvement in performance from 3 to 36 months postinjury. Although further investigation is needed to ascertain the usefulness of TQT in rehabilitation and special education settings, the cross-sectional and longitudinal findings provide support for clinical application of this procedure. It is conceivable that using three trials (Goldstein & Levin, 1991) instead of the single item version of the TQT employed in this study might be more sensitive to changes in performance over time.

We included the TOL because our previous principal components analysis of EF measures in head injured and normal control children disclosed that the TQT and these tests loaded on a common factor (Levin et al., 1996). The present cross-sectional study also revealed that the TOL was sensitive to severity of CHI and that the degree of impaired performance was disproportionate in young children relative to older children and adolescents. Multiple regression indicated that the volume of left frontal lesion incremented prediction of the percent of problems solved on Trial 1 and the initial planning time, whereas left nonfrontal lesions contributed to predicting the number of broken rules. This pattern of findings was generally corroborated in the longitudinal study, which also showed that the extent of improvement over 3 years tended to be greater in children who sustained severe head injury (e.g., number of broken rules). Plotting the longitudinal TOL data separately by complexity of the problems indicated the presence of a ceiling effect on solving the problems requiring two to three moves. In concert with the conclusions reached in a recent developmental study of the TOL (P. Anderson et al., 1996), we postulate that extending the level of difficulty would enhance the usefulness of the test.

The WCST was included in this study because of its extensive and long-standing use as a measure of EF. Although Levin et al. (1991) found a developmental trend for WCST variables, the present cross-sectional and longitudinal studies indicated that this test was not consistently sensitive to severity of CHI as measured by the GCS score. Consistent with some, but not all studies of adults with focal brain lesions (S. Anderson et al., 1991), multiple regression showed that left and right frontal lesions incremented prediction of the number of categories attained and the percent perseverative errors as compared to the level of prediction obtained by using the severity of CHI and age without considering the MRI findings. Although the WCST was sensitive to changes over 3 years, the extent of improvement was related to severity of CHI, but not age at injury. Further investigation of EFs in children is needed to isolate the components of working memory, attention, and other factors contributing to performance.

The relevance of EF measures to adaptive functioning is pertinent to their potential clinical application. To explore the ecologic validity of the three cognitive measures used in the cross-sectional study, we correlated the scores with the Vineland Adaptive Behavior Composite percentile, which was obtained from a structured interview with the parent on the day of assessment. Our finding of moderate, albeit consistent, correlations in the predicted direction provides support for the relevance of all three cognitive measures to adaptive functioning. Furthermore, a regression model on the EF measures showed the percent conceptual level responses on the WCST and the percent hypothesis questions on the TQT were the most efficient predictors for the Vineland Composite percentile, accounting for nearly 19% of the variance. However, we acknowledge that using a functional outcome measure such as school performance may yield correlations with the EF measures that differ from the present findings for the Vineland.

The clinical application of EF measures to the rehabilitation and special education of head injured children also merits investigation. Denney et al. (1979) showed that modeling techniques could modify the question-asking strategies of young children, a finding that invites replication in a special education or rehabilitation setting. Finally, studies addressing the role of emotional factors in EF performance are also relevant to clinical application, particularly in view of the purported relationship between prefrontal dysfunction and neuropsychiatric disturbance in children (Price et al., 1990).

Our findings raise three major measurement issues concerning the application of EF tests to assess the cognitive effects of CHI in children. First, the interaction of CHI severity with age was inconsistent across the three EF tests. Although severity of injury interacted with age on the TOL, this interaction was not found on the TQT or the WCST. Second, the separate interactions of age at study with severity group and level of complexity on the TOL are also open to competing interpretations. If injury at a young age has disproportionate effects on cerebral development (e.g., by disrupting myelination), these adverse effects should be present in both cross-sectional and longitudinal analyses. Although the interactions with age found on the TOL appeared to be stronger in the cross-sectional study than the longitudinal investigation, this divergence could be due to the differences in sample size or inclusion of normal controls in the cross-sectional study whose performance was more likely to have been characterized by ceiling effects. Ceiling effects in the performance by older children on the TOL might have also contributed to the Age × Severity interaction in the longitudinal study due to repeated exposure to the task. With the exception of severely injured children younger than age 9 years, ceiling effects appear to be present in the cross-sectional and longitudinal TOL data. Implications of our findings are that the range of difficulty should be increased for the TOL, and that a parallel form should be developed for repeated administration. Finally, the usefulness of these EF measures can be challenged on the basis that impaired performance following brain injury might be attributable to group differences in intellectual ability as reflected by standard psychometric tests or to a pervasive deficit in attention that adversely affects all cognitive measures. Although the design of this study does not allow us to address the specificity of EF deficit following pediatric brain injury, this issue merits further investigation in future studies.

ACKNOWLEDGMENTS

Supported by Grant NS-21889 to Harvey Levin, Ph.D. The authors wish to thank Dianne Mendelsohn, M.D., and Joel Yeakley, M.D., for assistance in the interpretation and acquisition of MRI scans, and Angela D. Williams for editorial and word processing assistance. The University of Texas Medical School Clinical Research Center provided support for the follow-up studies.

REFERENCES

Adams, J.H., Graham, D.I., Scott, G., Parker, L.S., & Doyle, D. (1980). Brain damage in non-missile head injury. *Journal of Clinical Pathology*, *33*, 1132–1145.

Anderson, P., Anderson, V., & Lajoie, G. (1996). The Tower of London test: Validation and standardization for pediatric populations. *Clinical Neuropsychologist*, *10*, 54–65.

Anderson, S.W., Damasio, H., Jones, R.D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology*, *13*, 909–922.

Cockburn, J. (1995). Performance on the Tower of London test after severe head injury. *Journal of the International Neuropsychological Society*, *1*, 537–544.

Damasio, H. (1991). Neuroanatomy of frontal lobe in vivo: A comment on methodology. In H.S. Levin, H.M. Eisenberg, & A.L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 92–124). New York: Oxford University Press.

Denney, D.R. & Denney, N.W. (1973). The use of classification for problem solving: A comparison of middle and old age. *Developmental Psychology 9*, 275–278.

Denney, N.W., Jones, F.W., & Krigel, S.H. (1979). Modifying the questioning strategies of young children and elderly adults with strategy-modeling techniques. *Human Development*, *22*, 23–36.

Fletcher, J.M., Ewing-Cobbs, L., Miner, M.E., Levin, H.S., & Eisenberg, H.M. (1990). Behavioral changes after closed head injury in children. *Journal of Consulting and Clinical Psychology*, *58*, 93–98.

Goldstein, F.C. & Levin, H.S. (1991). Question-asking strategies after severe closed head injury. *Brain and Cognition*, *17*, 23–30.

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

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual–Revised and expanded.* Odessa, FL: Psychological Assessment Resources.

Levin, H.S., Culhane, K.A., Hartmann, J., Evankovich, K., Mattson, A.J., Harward, H., Ringholz, G., Ewing-Cobbs, L., & Fletcher, J.M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology*, 7, 377–95.

Levin, H.S., Culhane, K.A., Mendelsohn, D., Lilly, M.A., Bruce, D., Fletcher, J.M., Chapman, S.B., Harward, H., & Eisenberg, H.M. (1993). Cognition in relation to magnetic resonance imaging in head-injured children and adolescents. *Archives of Neurology*, *50*, 897–905.

Levin, H.S., Fletcher, J.M., Kufera, J.A., Harward, H., Lilly, M.A., Mendelsohn, D., Bruce, D., & Eisenberg, H.M. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks of head-injured children and adolescents. *Developmental Neuropsychology*, *12*, 17–34.

Levin, H.S., Mendelsohn, D., Lilly, M.A., Fletcher, J.M., Culhane, K.A., Chapman, S.B., Harward, H., Kusnerik, L., Bruce, D., & Eisenberg, H.M. (1994). Tower of London performance in relation to magnetic resonance imaging following closed head injury in children. *Neuropsychology*, *8*, 171–179.

Mendelsohn, D., Levin, H.S., Bruce, D., Lilly, M.A., Harward, H., Culhane, K., & Eisenberg, H.M. (1992). Late MRI after head injury in children: Relationship to clinical features and outcome. *Child's Nervous System*, *8*, 445–452.

Mosher, F.A. & Hornsby, J.R. (1966). On asking questions. In J.S. Bruner, R.R. Olver, P.M. Greenfield, J.R. Hornsby, H.J. Kenney, M. Maccoby, N. Modiano, F.A. Mosher, D.R. Olson, M.C. Potter, L.C. Reich, & A.M. Sonstroem (Eds.), *Studies in cognitive growth* (Ch. 4, pp. 86–102). New York: John Wiley & Sons, Inc.

Price, B.H., Daffner, K.R., Stowe, R.M., & Mesulam, M.M. (1990). The comportmental learning disabilities of early frontal lobe damage. *Brain*, *113*, 1383–1393.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London*, *298*, 199–209.

Sparrow, S.S., Balla, D.A., & Cicchetti, D.V. (1984). *Vineland Adaptive Behavior Scales* (Interview ed., Survey Form). Circle Pines, MN: American Guidance Service.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: A practical scale. *Lancet*, *2(7872)*, 81–84.