4.0 NORMATIVE SAMPLE

4.1 METHOD

Subjects

The normative sample consists of 183 boys, approximately 60 in each of the 7-8, 9-10, and 11-12 year age ranges. The sample was drawn from four government schools in Sydney, Australia. The majority of subjects (76%) came from one school where a *seriatim* sample was taken from all those whose parents gave permission to be involved in the study. The remaining 24% were randomly selected from the other three schools. In total, 89% of parents gave permission for their child to be involved in the study. Eighty-seven percent of the 7-8 year olds in the study came from the one school as did eighty-eight percent of 9-10 year olds and fifty percent of 11-12 year olds. The four schools were from the same geographic area, which, according to the 1991 Australian Census, was comparable to the Sydney City population in terms of home ownership, income, language spoken at home, ethnicity, and educational level. This information is detailed in Table 9. It is evident that the percentage of residents from both population groups who were born in Australia, who speak English at home and who own a home, are comparable. In addition, the Annual Family Income of the two populations was strikingly similar, with a mean difference in percentages for each of the five income ranges being only 3.2%. There was some variation between the two populations on "occupation", detailed in Table 9, although this was not considered a significant difference for the purposes of establishing that the Normative Sample was representative of the Sydney City Sample. In summary, the data indicated that the Normative sample would be considered as coming from a typical Sydney home.

Table 9: Demographic comparison between normative sample and Sydney City sample based on figures from the 1991 Australian Bureau of Census Data.

Category	Normative Sample	Sydney City Sample
Born in Australia	67.8%	66.9%
English spoken at home	71%	72.2%
Highest educational qualification achieved by parent	Tertiary Degree: 6.7% Skilled vocation: 15.7% Not qualified: 64.4% Not Stated: 13.2%	13.4% 15.6% 58% 13%
Occupation	White collar: 20.6% Blue collar: 58% Unskilled: 16.8% Not Stated: 4.6%	32% 50.6% 12.3% 5.1%
Annual Family Income	0-$20,000:15.1% -$40.000:24% -$50,000:13.8% -$60,000:10.4% -$100,000:13.8% Not stated: 22.9%	16.3% 21.8% 11.9% 9.6% 20.3% 20.1%
Home Ownership	56.5%	52.4%

The children were screened for evidence of neurological impairment (serious birth difficulties, loss of consciouness, epilepsy and head injury), and those with evidence of impairment were removed from the sample. Other risk factors could have been applied. For example, children with significant auditory, linguistic or visual deficits as well as children with diagnosed attention deficit disorder. However, it was considered that those factors employed in this study represented an adequate gross screening procedure. Twelve percent, or 26 of those sampled, were ruled out on the basis of this risk for CNS impairment. See Appendix 6 for the screening *proforma* completed by parents. Table 10 shows the distribution of subjects by age.

Table 10: Characteristics of Normative Sample

Age (yrs)	n	Age in mths
7	31	91.48
8	31	100.94
9	30	113.21
10	30	125.43
11	31	138.32
12	30	147.10

Instruments

The tests were administered in the same order for every subject and given in two testing sessions, each of approximately one to one and a half hour's duration.

4.1.1. SEALS

The Seals task is based on the Tower of Hanoi, a developmental measure used in problem-solving and described previously. The task was developed in the 'active memory' mode after Klahr and Robinson (1981), in that subjects were instructed to verbalise their moves to reach the goal-state without the benefit of actually physically executing the moves.

Once the subjects understood the task through their responses on the two-ball problem, no feedback was given on the efficacy of their responses during the three and four-ball problems. Three problem types are presented. In the two-ball problem, the seal on the left of the screen balances two balls, a small green ball resting on top of a medium-sized red ball. In an inset at the top of the screen is displayed the goal-state which consists of the same balls, in order, resting on the seal on the right of the screen. The three and four ball problems are similarly presented, with the green, red and larger blue ball balancing for the three ball problem, and the green, red, blue and yellow balls balancing for the four-ball problem.

The task was programmed so that it began with all the balls balancing on the Seal to the left of the screen, and with each successive error moved one step closer to the goal-state. Subjects have to successfully solve two successive trials for each problem type before moving on to the next problem level.

The following instructions appear on the screen:

"There are three seals, one of which is balancing coloured balls on its nose. You have to tell me how you would move the balls, one at a time, so that they finish in the same position as the seal in the box at the top of the screen."

The researcher reads the instructions to the subject, and adds:

"There are three rules you have to remember:

* You can only move one ball at a time
* You cannot place a larger ball on top of a smaller ball...."

The next screen shows a seal balancing a large ball on top of a small ball, with a large cross appearing over the balls, indicating an illegal move.

* You have to solve the problem in the smallest number of moves.

All responses are recorded by the researcher on a specially prepared proforma setting out all the possible state spaces for each problem type. (See Appendix 7 for examples of these proformas). The computer screen is reproduced in Figure 12.

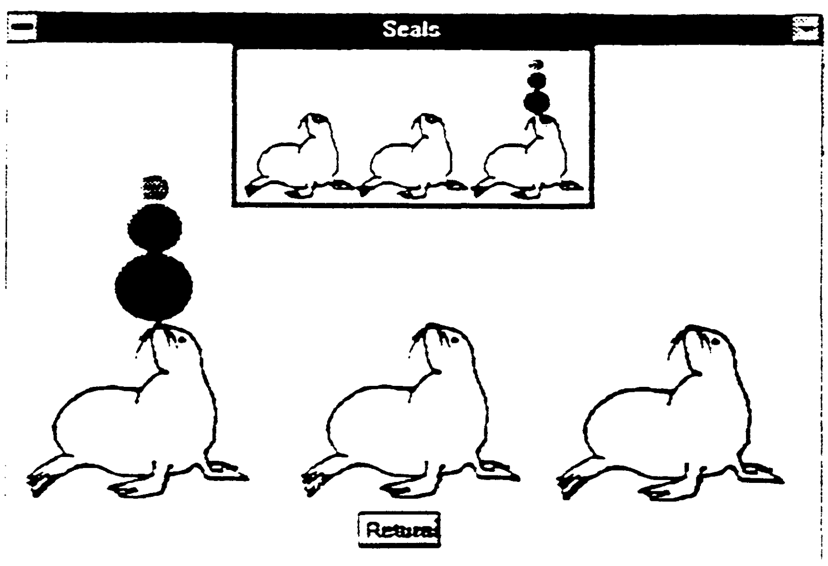


Figure 12: Example of screen for Seals Task.

Two-ball problem

During this problem type, verbal feedback is given as soon as a rule violation is detected. Rule violations are brought to the subject's notice and the correct rule stated. The next screen is then presented with the balls moving one step closer toward the goal-state.

However, moves which are not of the optimum type, but which do not violate one of the three rules, (that is, does not move more than one ball at a time and does not place a larger ball on top of a smaller one) are allowed to continue until either the subject breaks a rule, indicates that he or she cannot continue, or completes the task. Self-corrections are permissible, and, in all likelihood, a subject may realise half-way through the task, that an incorrect sequence was followed.

It is only during this two-ball problem that subjects watch the balls animated in their moves toward the goal state.

Three-ball and Four-ball problem

No feedback on performance is given during these problem types. As soon as a rule violation or error occurs, the experimenter presses the <Enter> key on the keyboard, and the next screen appears with the balls moved one step toward the goal-state and subjects are asked to solve that problem starting from that position.

As in the two-ball problem, moves which are consistent with the rules are allowed to continue. That is, provided the subject moves one ball at a time and does not place a larger ball on top of a smaller ball, he or she is allowed to continue with the task. Self-corrections are permissible.

The researcher records all responses by the subject, including self-corrections, queries regarding rules or possible moves, restating of rules and other comments. Scoring is based on Borys, Spitz and Dorans (1982) where six trials are given for each problem type and criterion for completion is the correct solution in the minimum number of moves over two successive trials.

4.1.2. BALANCE BEAM

The following instructions appear over three screens. The researcher reads each screen to the subject, while pointing. The researcher presses the <Enter> key on the computer keyboard to move on to the next screen. The text changes with each screen while the balance beam remains on the screen:

Screen 1: *This is a balance beam with metal disks placed on these pegs. All the disks weigh the same.*

Screen 2: *What would happen if the bar were to swing free? Would the left side go down, would the right side go down, or would it stay the same as it is now? Use the mouse to click on the one picture at the top of the screen which indicates your answer.*

Screen 3: *After you have clicked on a picture, the balance beam will move and you will see which way it swings. The next problem will then appear on the screen. Click on a picture now.*

At the top of the screen are three icons showing a small balance beam in the horizontal position, a small beam with the left hand side dropped down, and one showing the right hand side dropped down. Subjects respond by clicking on one of these icons indicating the outcome.

The task has been programmed to maintain running records of students' results which are saved to the hard disk drive for later retrieval.

When subjects click on the icon at the top of the screen, the balance beam then moves according to gravity. Subjects obtain immediate feedback on the accuracy of their responses and can make necessary adjustments based on this information. An example of a screen is presented in Figure 13:

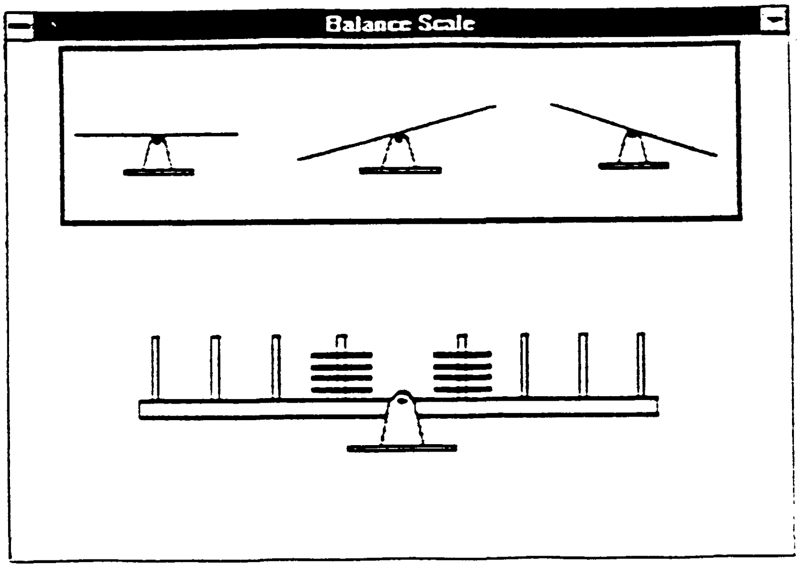


Figure 13. Example of a Balance Beam screen.

4.1.3. FISH TASK

The following instructions appear on the screen The experimenter reads out loud while pointing to each word:

Fish will appear one at a time on the screen. When you see a fish which is not the same as the one which is now on the screen (picture of Background fish appears on screen), click on the mouse. When you are correct, a bell will sound.

The task consists of thirty-five drawings of fish presented individually on the computer screen. There are twenty of the same Background fish and fifteen different Target fish. The computer records correct responses and stores these onto the hard disk drive for later retrieval.

4.1.4. PIGGY BANKS

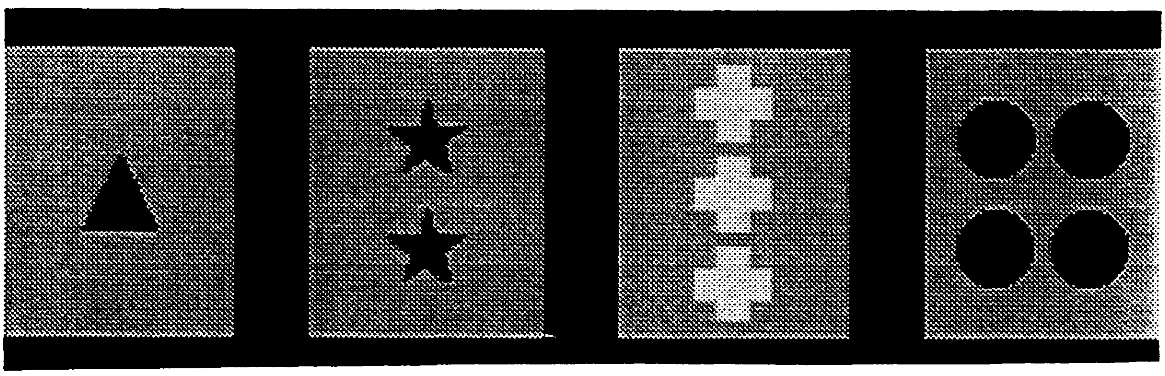
The following instructions appear on screen when the program opens:

A coin will be placed in one of the piggy banks. Watch carefully and remember which piggy bank has the coin. Take the mouse and click on the piggy bank which contains the coin. If you are right the coin will appear and a bell will sound.

The program records correct responses which are stored onto the hard disk drive for later retrieval.

4.1.5. WISCONSIN CARD SORTING TEST (WCST)

Lau and Perdices (1990), using the Nelson (1976) adaptation of the WCST developed a computerised version of the WCST involving the twenty-four cards presented twice. The first computer screen of the program contains four cards with geometric figures on the top half of the screen with written instructions on the bottom half of the screen. The experimenter reads the instructions out loud to the subjects and presses the <Enter> key on the keyboard to move to the next instruction screen. During the four instruction screens, the top half of the screen with the four cards remains constant.



Screen 1: *In this test, the four cards that are displayed in the upper half of the screen are called SAMPLE CARDS.*

TEST CARDS will be shown to you, one at a time, at the bottom of the screen. You must match the TEST CARDS with one of the above SAMPLE CARDS.

Screen 2: *When you have decided which SAMPLE CARD you want to match the TEST CARD with, press the number (1, 2, 3, 4) that corresponds to that SAMPLE CARD.*

For example, if you decided to match the TEST CARD with SAMPLE CARD #1, press 1, etc.

Screen 3: *It is up to you to decide how to match the TEST CARD with the SAMPLE CARDS. After each match, a "CORRECT" or "INCORRECT" message will appear on the screen.*

You must use this information to work out how to match the TEST CARD appropriately, so that you get as many correct matches as possible.

Screen 4: *Do you have any questions? When ready, press <ENTER> to begin the Test.*

When the first TEST CARD appears on the lower half of the screen, the experimenter reinforces the written instructions by saying the following: "You have to decide which one of these (pointing) cards at the top of the screen best matches this TEST CARD (pointing). When you have decided, press the number that is on top of the SAMPLE CARD (pointing to the numbers on top of the cards and the numbers on the keyboard). When you have made your choice, a "Correct" sign (with a green background) will tell you that you were right, or an "Incorrect" sign (with a red background) will tell you that you were wrong. You have to try to get as many correct choices as you can."

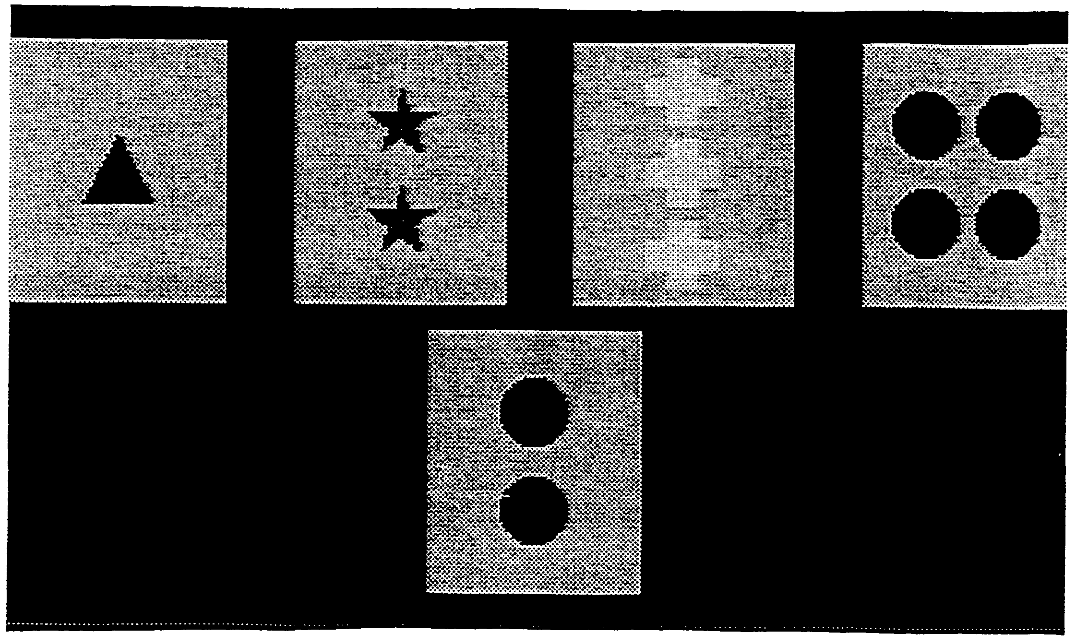


Figure 14: Example of computer screen for WCST (Lau and Perdices, 1990)

After six consecutive correct sorts, the criterion automatically changes.

Subject's responses were recorded by the computer onto a separate file on the hard disk drive for later retrieval. The computer recorded time taken to complete the test, number of sets of six consecutive sorts, number of perseverative errors, and total number of errors. Appendix 8 contains an example of the computer printout of a subject's scores on the WCST.

A perseverative error was defined as the same incorrect sorting criterion that was used following the first time that it had been demonstrated to be incorrect.

4.1.6. AUSTIN MAZE

This is an adaptation of the "stepping-stone maze" of Barker (1931). Developed by Walsh (1985), it studies self-corrective behaviour as well as the ability to follow rules. According to Lezak (1983), the Austin Maze requires the subject to learn a long pathway through a 10 x 10 grid of squares that momentarily light up - green for correct steps and red for incorrect - along the pathway.

A computerised version of the Austin Maze has been developed by Lau and Perdices (1990). Subjects are given a simple practice trial to familiarise themselves with the test and to demonstrate both legal and illegal moves. Learning criterion is two consecutive error-free trials, with a maximum of twenty trials presented. The Milner (1965) pathway version of this test was used. The computer records the following information which is stored on the hard disk drive:

- time taken to complete each trial,
- total number of moves made in each trial,
- coding and frequency of five error types.

The five error types are as follows:

Error Type 1:	Does not go back to the last correct square
Error Type 2:	Back tracks
Error Type 3:	Perseverative error with respect to the last trial
Error Type 4:	Diagonal moves
Error Type 5:	Perseverative error within the trial.

The following instructions, adapted from Walsh (1985, 1991), were given during a Practice Example.

This is a maze and you have to find the hidden pathway to move from here (point to "begin") *to the end* (point to "end") *on the screen. Your task is to find this pathway by moving one step at a time across the screen and then to see how much of it you can learn by going across a number of times.*

There are several rules in this game:

(1) You can only move in the following ways (demonstrate using keyboard) *left* (press — key), *right* (press _ key), *down* (press key) *and up* (press key). *You cannot move diagonally.*

(2) If you move to the correct square, the square will light up green (demonstrate), *and you are free to keep going. If you move to a square which is not on the pathway it will light up red* (demonstrate), *and a bell will sound. You must go back one step to your last correct choice and try another one.*

(3) If you move to a square which lights up green with a square inside, it means that you have gone back over a square which was already correct. You must go back to your last correct choice and try again.

We are going to do a practice example so that you know what you have to do.

The subject is given assistance to complete the practice example. Following this practice example, the instructions for Trial 1 of the test are predicated by:

This is a new maze. I want you to find the hidden pathway by making as few mistakes as you can, using the same rules that you used in the practice example. Go ahead.

For subsequent trials, the researcher says:

This is the same maze that you just did. Try to remember it and make as few mistakes as you can.

The researcher gives encouragement to the child to keep going and make as few errors as possible. Appendix 9 contains an example of a computer printout of a subject's results on the Austin Maze.

4.1.7. AUDITORY VERBAL LEARNING TEST (AVLT)

The AVLT (Rey, 1964) has been demonstrated to measure three stages of verbal memory: registration, storage and retrieval (Vakil and Sheleff, 1990). According to Lezak (1983), it measures immediate memory span, retroactive and proactive interference, and retention following an interpolated activity. The test consists of three sections:

* Learning-recall trials (trials 1-7)
* A thirty minute delay
* Delayed recall test (trial 8) and Recognition test (trial 9).

The following word lists, taken from Lezak (1983) were used for the AVLT:

List A	**List B**
Drum	Desk
Curtain	Ranger
Bell	Bird
Coffee	Shoe
School	Stove
Parent	Mountain
Moon	Glasses

Garden	Towel
Hat	Cloud
Farmer	Boat
Nose	Lamb
Turkey	Gun
Colour	Pencil
House	Church
River	Fish

Procedure:

For Trial 1, List A is read by the examiner at the rate of one word per second after giving the following instructions:

I am going to read a list of words. Listen carefully, for when I stop you are to say back as many words as you can remember. It doesn't matter in what order you repeat them. Just try to remember as many as you can.

For Trials 2, 3, 4 and 5 the Examiner rereads List A with the following instructions:

*Now I'm going to read the same list again, and once again when I stop I want you to tell me as many words as you can remember, **including words you said the last time.** It doesn't matter in what order you say them. Just say as many words as you can remember whether or not you said them before.*

For Trial 6, the following instructions are given:

Now I'm going to read a second list of words. This time, again, you are to say back as many words of this second list as you can remember. Again, the order in which you say the words does not matter. Just try to remember as many as you can.

List B is then read by the Examiner and the subject recalls as many words as he can.

Trial 7 is an immediate recall of words from List A without presentation. Trial 8 is a delayed-recall test, without presentation, of the words from List A, and following a delay of thirty minutes.

Trial 9, which immediately follows Trial 8, is a Delayed Recognition Test (Geffen, Forrester, and Tuckfield, 1992). The subject is given a form containing fifty words, consisting of the fifteen words from each of the recall lists, as well as twenty distractor words with associated meaning (semantic distractors), or similar sound (phonemic distractors), or both (semantic-phonemic distractors), to a word from one of the lists. The subject ticks those words which he recognises from either list, and indicates whether each word came from List A or List B.

The following instructions are given for Trial 9:

On this page there are some words that were in the lists that you learnt and some that were not. I want you to place a tick next to the words that you learnt before. If you can remember which list the words come from, put '1' next to words from the first list and '2' next to words from the second list.

See Appendix 10 for Trial 9 Delayed Recognition Form.

For Trials 1-8, each word is written by the tester as recalled. These are entered onto an IBM PC using a Flinders Auditory Word Memory Program designed for scoring the AVLT (Forrester and Geffen, 1991). The program generates data on **acquisition rate** (Trials 1-5), **proactive interference** (Trial 6/Trial 1), **retroactive interference** (Trial 7/Trial 5), **forgetting** (Trial 8/Trial 7), and **retrieval efficiency** (prop. Trial 8/P(A) List D).

Scores recorded for each list are:

Words correctly recognised (in each list).
correctly assigned
incorrectly assigned
not assigned
False positives
semantic associations
phonemic associations
semantic-phonemic associations

4.1.8. OTIS-LENNON SCHOOL ABILITY TEST (OLSAT)

The Otis-Lennon School Ability Test, Sixth Edition, (Otis and Lennon, 1989) is designed to measure verbal and non-verbal reasoning skills that are most closely related to scholastic achievement. OLSAT contains separate test levels for each grade from Kindergarten through to Year 3. Years 4 and 5 do the same test level, and Years 6-8 do the same test. The classification of items as Verbal and Nonverbal hinges upon whether knowledge of the English language is requisite to answering the items. Raw scores are firstly converted to a scaled score and then translated into a School Ability Index which is a normalised standard score with a mean of 100 and a standard deviation of 16. Test administration varies with each grade level, and separate instructions and procedures were used with each grade. A contents outline for each separate test is detailed below (Otis and Lennon, 1989, p 11):

Cluster/Item Type	C Grade 2	D Grade 3	E Grades 4-5	F Grades 6-8
VERBAL				
VerbalComprehension				
Following Directions	12			
Antonyms		5	4	4
Sentence Completion		4	4	4
Sentence Arrangement		4	4	4
Verbal Reasoning				
Aural Reasoning	12			
Arithmetic Reasoning	6	4	4	4
Logical Selection		3	4	4
Word/Letter Matrix		4	4	4
Verbal Analogies		4	4	4
Verbal Classification		4	4	4
Inference			4	4
NONVERBAL Pictorial Reasoning				
Picture Classification	6			
Picture Analogies	3			
Figural Reasoning				
Figural Classification	6	8		
Figural Analogies	9	9	6	6
Pattern Matrix	3	4	6	6
Figural Series	3	4	6	6
Quantitative Reasoning				
Number Series		6	6	6
Numeric Inference		6	6	6
Number Matrix		6	6	6

It is evident that the OLSAT measures quite different aspects of school ability for each grade which is tested. This is particularly so for the younger age group where the emphasis is on aural reasoning, the ability to follow verbal directions, picture classification and picture analogies The items in these classifications do not appear in the later age groups. In the present study, the OLSAT was used to test the sensitivity of a measure of academic ability to differentiating the performance of children with brain injuries. Recall that previous studies had indicated no significant difference on measures of school ability following a significant brain injury (Williams and Mateer, 1992).

4.2 PROCEDURE

The Tests were administered in two separate sessions, two days apart. The order of presentation of tests for all subjects was:

Session One:

Auditory Verbal Learning Test (Trials 1-7)
Seals Test
Fish Test
Piggy Bank
Auditory Verbal Learning Test (following a delay of 30 minutes, Trials 8 and 9).
Balance Beam
Austin Maze
Wisconsin Card Sorting Test.

Session Two:

Otis-Lennon School Ability Test (OLSAT).

5.0 RESULTS

5.1 Normative Sample Outliers

Three subjects, from the pool of 188 subjects tested, were eliminated from the sample because their results during the Fish task were considered invalid. In each case, the subjects realised part way through the exercise, that they had not listened to the instructions and were responding to the Background rather than the newly presented Target fish as it appeared on screen. Those eliminated were Subject #114, aged 8-1years with a score of 14%, Subject #234, aged 9-3years with a score of 2%, and Subject #142, aged 11-6years, with a score of 8%.

An analysis was conducted on raw scores for the four measures of Seals, Fish, Balance Beam and Piggy Bank to determine the outliers on the remaining 185 subjects. Within each age range the subjects who had at least two scores greater than -1.5 standard deviations below the mean were identified. These are detailed in Table 11:

Table 11: Outliers scoring greater than -1.5 Sd below mean scores on experimental developmental tests

Age Range	Subject	Seals	Fish	Balance	Piggy
Junior (7-8 yr)	#130	-2.5 SD	-1.5 SD	-2.75 SD	-
	#111	-1.6 SD	-1.95 SD	-	-
Middle (9-10 yrs)	#045	-1.7 SD	-	-3 SD	-
	#59	-3 SD	-	-	-1.6 SD
	#19	-1.9 SD	-1.8 SD	-3.9 SD	-1.6 SD
Senior (11-12 yrs)	#304	-2 SD	-4.3 SD	-3.59	-

All normative subjects had been screened for possible neurological involvement. However, the six subjects in Table 11, when compared with their aged peers in this sample, scored significantly below the mean in a number of the tests. Further enquiries as to the subjects' backgrounds were made from school records and discussions with school personnel in order to gain an understanding of these results.

Two interesting results emerged. Subject #19, the only subject to score greater than -1.5 SD below the mean on the four tests, had in fact suffered from convulsions when he was six months of age. This fact was not recorded by his parent on the screening sheet, but emerged in follow-up discussions.

Subject #130 who scored greater than -1.5 SD below the mean for three of the four tests also had a negative parent report to possible neurological involvement. On further investigation with the family, it was revealed that the mother suffered epileptic seizures during pregnancy. She had not indicated this on the original parent proforma. The seizures may have resulted in anoxia to the foetus causing neurological impairment (Walsh, 1991).

Nothing remarkable was discovered about the remaining four subjects. However, it is considered that there may have been neurological involvement with these subjects, perhaps at the time of birth, despite the lack of any obvious positive signs. There is evidence, for example, that 4% of 'healthy' infants with normal ultrasounds at discharge from hospital, were found to have developed major neurodevelopmental impairments when assessed at age 8 years (Roth, Baudin, McCormick, Edwards, Townsend, Stewart and Reynolds, 1993). The two outliers with neurological history signs from parent follow-up were consequently excluded from the study. The other four outliers were retained in the sample. This left a total of 183 subjects which were included in the sample. The three age groups contained 62 in the Junior Group (7-8 years), 60 in the Middle Group (9-10 years), and 61 in the Senior Group (11-12 years). For a full listing of the raw scores, separated by age groups, see Appendix 11.

5.2 FISH TASK

Two scores were calculated from subjects' raw scores on this task: A Total Score (percent correct), and Perseverative Errors. Perseverative Errors were calculated from the second consecutive error in each separate string of errors. If a subject made three consecutive errors, then the perseverative error score was two for that group of errors. These perseverative errors were added, divided by the total number of errors and converted to a percentage score.

Mean scores and standard deviations, in brackets, for these measures are set out in Table 12:

Table 12: Results of Fish Task

Age Group	Total Score (%)	Perseverative Errors
Junior (7-8yrs, n=62)	88.6 (15.25)	29.5 (20.7)
Middle (9-10yrs, n=60)	91.8 (11.4)	33.1 (21.5)
Senior (11-12 yrs, n=61)	96.2 (6)	21.6 (22)

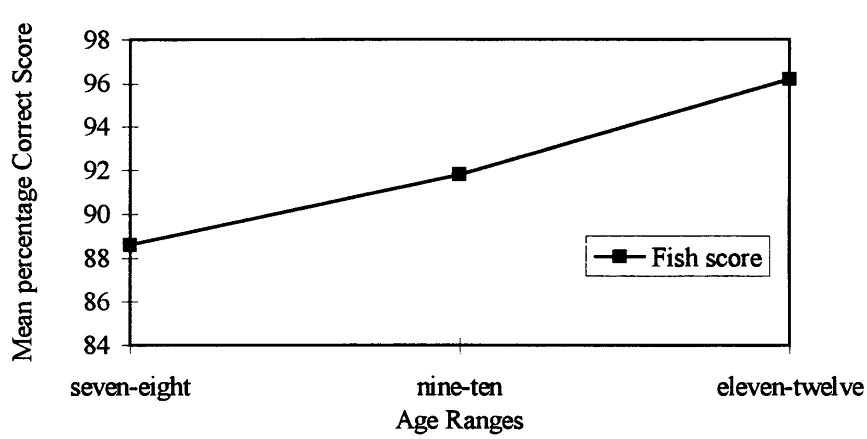


Figure 15: Comparison of three age groups on Fish Scores

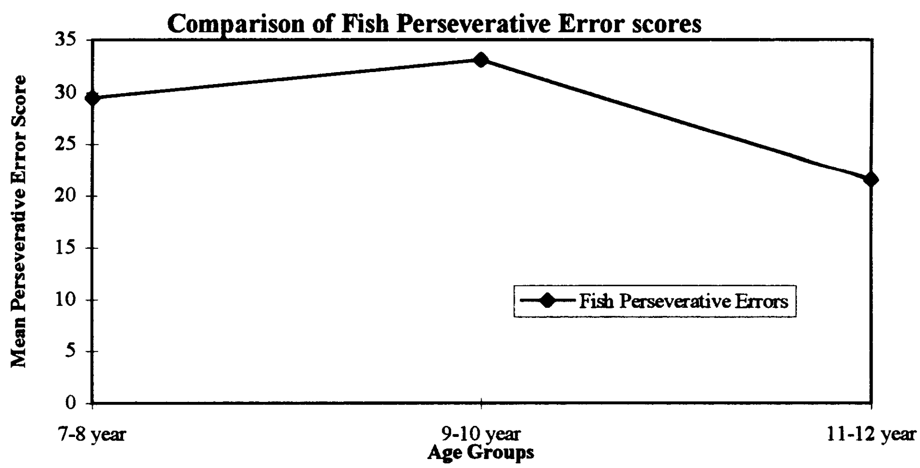


Figure 16: Comparison of Fish Perseverative Error scores across three age groups

On first inspection, the results for the total score indicate a developmental trend with increasing accuracy with age. At this point in the analysis, the commonly recommended practice is to conduct a MANOVA, and if the null hypothesis is rejected, conduct follow-up tests to clarify the nature of the differences between groups (Olson, 1979). This procedure is based on the assumption that the only function of the initial overall test is to protect the second stage of analysis from inflation of the Type 1 error rate (Bird and Hadzi-Pavlovic, 1983). However, there is evidence that the Bonferroni critical values approach provides a more powerful test of paired comparisons than tests of multivariate analysis (Harris, 1985). According to Harris (1985) and Bird (1975), the Bonferroni approach provides conservative tests of statistical significance, but it nevertheless often yields less stringent critical values (and thus more powerful tests), especially when the total number of comparisons is small. For these reasons, the Bonferroni adjustment for paired comparisons was applied to the data.

Only T-tests for independent samples using Bonferroni adjustments (Miller, 1966), were computed using Testat. The Bonferroni adjustment was alpha/3 as there were 3 tests computed for paired-comparisons: Junior-Middle, Junior-Senior and Middle-Senior. For the p-value, significance was therefore set at $p<0.017$ for the Bonferroni-adjusted value.

Table 13: T-tests for Fish Total Score

Age Groups	T-Score	D.F.	p-value
Junior-Middle	1.302	120	0.196
Junior-Senior	-3.625	121	0.000**
Middle-Senior	-2.665	119	0.009**

**significant at the Bonferroni-adjusted level

There is a significant developmental difference in Fish scores between the 9-10 year age group and the 11-12 year age group and between the 7-8 and 11-12 year age group as detailed in Table 13. This difference is not seen between the 7-8 year age group and the 9-10 year age group.

Differences in results on the Fish Perseverative Error scores were also tested through independent sample T-tests using Bonferroni probability adjustments. These are displayed in Table 14:

Table 14: T-tests for Fish Perseverative Errors

Age Group	T-score	D.F.	p-value
Junior-Middle	-1.116	59	0.269
Junior-Senior	1.762	60	0.083
Middle-Senior	2.769	59	0.003**

**significant at Bonferroni-adjusted level

Perseverative Errors scores do not follow a clear developmental trend. There is a significant difference in scores between the Middle age group and the Senior age group.

The Cronbach Alpha internal reliability score was computed using Testat. Reliability across all items was 0.886.

The Fish data were analysed using a one-parameter Rasch probability model (Rasch, 1960,1966). Through this procedure, both person and item parameters can be placed on the same linear scale and a goodness-of-fit for each item to the total Fish scale can be estimated. The Rasch model is scaled to have a mean of zero and a standard deviation of one. During the iterative process, the Item response routine excludes subjects with a zero or perfect score. Likewise, any trial which is responded to in exactly the same way by all subjects is excluded from the analysis. Rasch scores approaching zero indicate that the actual scores are close to the predicted scores for that particular sample. Large negative Item Difficulty scores are indicative of items which have excessively orderly responses (Wright and Masters, 1982). High scoring subjects obtain high scores and low scoring subjects obtain low scores. At face value it would appear that these items are good discriminators. However, according to Masters (1985), when items are unusually orderly (negative item difficulty score >-3), this inevitably masks some instability in the item. On the other hand, items with a large positive Rasch difficulty score (>3) are indicative of items where low scoring subjects have a high score and high scoring subjects have a low score (Wright and Masters, 1982). Items with large positive Rasch statistics are therefore poor discriminators and should be examined with a view to elimination or modification in further test development.

Mean scores for each item and Rasch item difficulties based on 126 useable subjects are listed in Appendix 12.

Trials #1, #9, #10, #16, #18, #21, #23, #26, #30, #31 and #35 are performing poorly in that they all have Rasch item difficulty scores larger than -3.00. This indicates they were quite easy items and a poor fit to the expected model. This is confirmed by examining mean scores for these items across the 126 useable cases. All eleven items had mean scores approaching one, indicating that the majority of subjects were correct in their responses to these items, and that the eleven items had low discriminatory power.

Eight of the eleven easiest items were Target fish which immediately followed a Background fish. By contrast, the three most difficult items, those that had the lowest mean score and Rasch item difficulty scores of one standard deviation or below the mean, were all responses to a Background fish which followed a Target fish - the reverse situation.

A possible explanation for this outcome is that subjects inhibited responses to the Background fish by not clicking on the mouse, and when a Target fish immediately followed were already in response inhibition mode. They presumably held back their response until confirmation of the stimulus and the appropriate response. Subjects are given two seconds in which to respond and this would have been ample time for them to click on the mouse. In the reverse situation, subjects had just clicked on a Target fish, were in "go" mode and continued to click on presentation of a Background fish. Reaction time to each stimulus would have confirmed this hypothesis. Unfortunately, it was not measured.

5.3 PIGGY BANKS

Four measures were derived from results on the Piggy Banks task. These were: Total Score (Percent Correct), Error type 1, Error type 2, and Error type 3. The three separate error scores relate to the three delay periods of 15 seconds, 20 seconds and 25 seconds. These scores were derived by summing the errors made on those trials with similar delay periods (15 second delay = Trials #1, #2, #3, #4; 20 second delay = Trials #5, #6, #7, #8; 25 second delay = Trials #9, #10), and dividing by the total error score. To avoid a zero, 1 was added to the numerator and denominator. Scores could therefore range between 0.09 and 1.

Paired comparison T-tests using a Bonferroni adjustment on the three Piggy Bank Error indices indicated that only one relationship, that of Pigerror2 and Pigerror3 in the Junior age group, had a significantly different score. In terms of the total differences between all the error scores, this one result was considered negligible, and it was decided to combine all error scores into one index.

The mean scores and standard deviations, in brackets, for the Total Score and Total Error Score for each age group are outlined in Table 15:

Table 15: Results for Piggy Bank

Age Group	Total Score (% correct)	Pigerror (Total)
Junior (7-8 yrs)	87.6 (13.9)	1.226 (0.264)
Middle (9-10yrs)	92.2 (9.15)	0.8 (0.261)
Senior (11-12yrs)	94.9 (7.55)	0.508 (0.229)

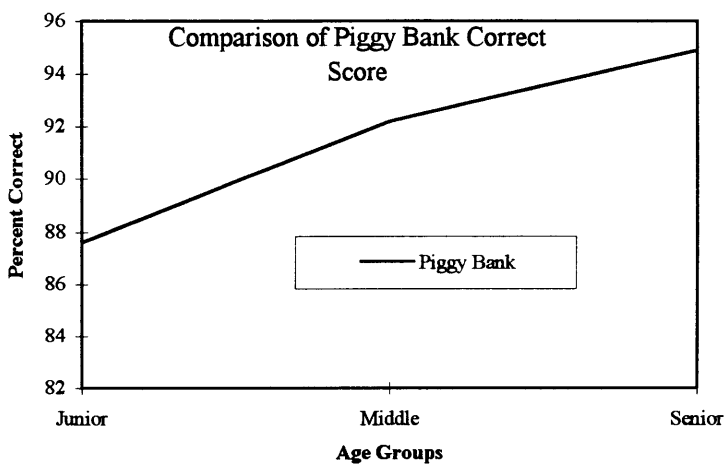


Figure 17: Comparison of Three Age Groups for Piggy Bank Correct Score

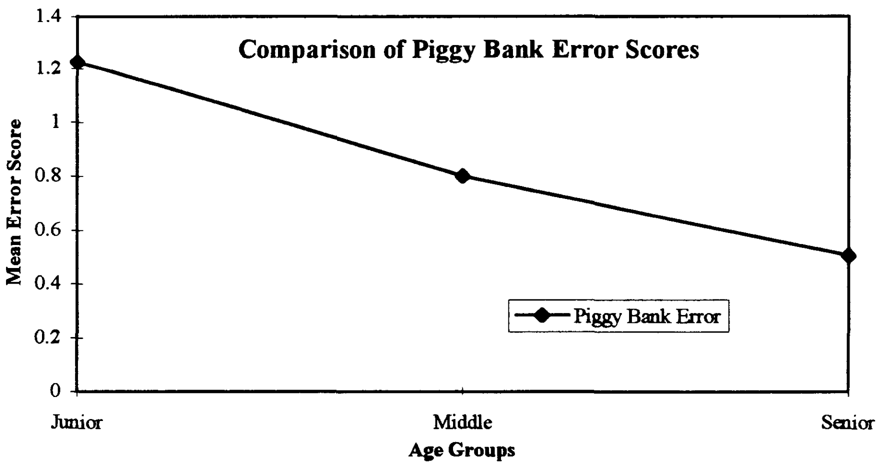


Figure 18: Comparison of Piggy Bank Error Scores across three age groups

Independent sample t-tests of significance with Bonferroni adjustments are outlined in Table 16:

Table 16: Piggy Bank Total Score Across Three Age Groups

Group	T-Score	D.F.	p-value
Junior-Middle	2.130	120	0.035
Junior-Senior	-3.519	121	0.001**
Middle-Senior	-1.680	119	0.096

** significant at Bonferroni adjusted level.

There was no clear incremental developmental progression across the age groups. A *post hoc* comparison of the mean scores for the three age groups suggests a small incremental improvement with age, although this result is not significant at the $p< 0.017$ Bonferroni adjusted confidence level. Notwithstanding, the difference between the 7-8 year age group and the 11-12 year group was significant.

The results of the Piggy Bank Error score indicate a lack of developmental progression, suggesting that this measure does not discriminate between the three age groups. There is a significant difference, however, between the Junior and Senior age groups.

Table 17: T-tests for Piggy Bank Total Error Scores

Age Groups	T-Score	D.F.	p-value
Junior-Middle	-1.886	119	0.064
Junior-Senior	-3.995	120	0.000**
Middle-Senior	-0.192	118	0.084

** significant at Bonferroni adjusted level

Cronbach alpha internal reliability scores across all 183 subjects was 0.392. The reliability indices for the Piggy Banks test are quite low. The individual item reliability scores are contained within a narrow band and are consistently poor. The individual item means, standard deviations and reliability scores are detailed in Appendix 13. An item response Rasch analysis using 97 useable cases (deleting subjects with perfect scores) was conducted using the Testat module of Systat. Table 18 contains the results:

Table 18: Rasch Item difficulty for Piggy Bank

Trial	Mean Score	Difficulty Score
1	0.835	-2.722
2	0.918	-4.052
3	0.918	-4.052
4	0.804	-2.365
5	0.918	-4.052
6	0.887	-3.451
7	0.763	-1.953
8	0.722	-1.588
9	0.835	-2.722
10	0.804	-2.365

From the Item Difficulty scores it appears that Trials #2, #3, #5 and #6 are poor discriminators in that they have Rasch difficulty scores greater than -3 or three standard deviations below the estimated Rasch model. This result indicates that those items were responded to in the same way by the majority of subjects with high scoring subjects consistently scoring highly on these particular items, and low scoring subjects consistently scoring poorly. It is concluded that the four items have low discriminatory power (Wright and Masters, 1982). The trials are a particularly poor fit to the estimated Rasch unidimensional model and would need to be deleted or modified in any further test development.

5.4 BALANCE BEAM

Two measures were derived from subjects' raw scores, a total percent correct, and a measure of perseverative errors. The perseverative errors were calculated from the second error in any consecutive combination of errors. These were then totalled and converted to a percentage of total errors. The mean results for each age group are set out in Table 19, with standard deviations in brackets.

Table 19: Mean scores on Balance Beam

Age Group	Total Score (% correct)	Perseverative Errors
Junior	62.9 (9.65)	60.177 (18.434)
Middle	67.5 (10.45)	54.367 (15.013)
Senior	70.7 (6.55)	49.820 (13.039)

Independent sample T-tests using Bonferroni adjustments are detailed below:

Table 20: T-tests on Balance Beam Total Score

Age Group	T-Score	D.F.	p-value
Junior-Middle	2.517	120	0.013**
Junior-Senior	-5.141	121	0.000**
Middle-Senior	-1.926	119	0.049

** significant at Bonferroni-adjusted level

The results of the t-tests indicate a progression across two age groups with a ceiling at the 9-10 year level. There is a trend for improvement in results between the 9-10 year group and the 11-12 year group although this difference is not significant at the $p < 0.017$ confidence level. The results also indicate an increase in variability in performance for the 9-10 year old Middle group. The standard deviations of Balance Beam Correct scores, outlined in Table 19 and Figure 20, indicate that the 9-

10 year old group were less stable in their responses than the other two age groups as evidenced by the fact that they had the largest standard deviation.

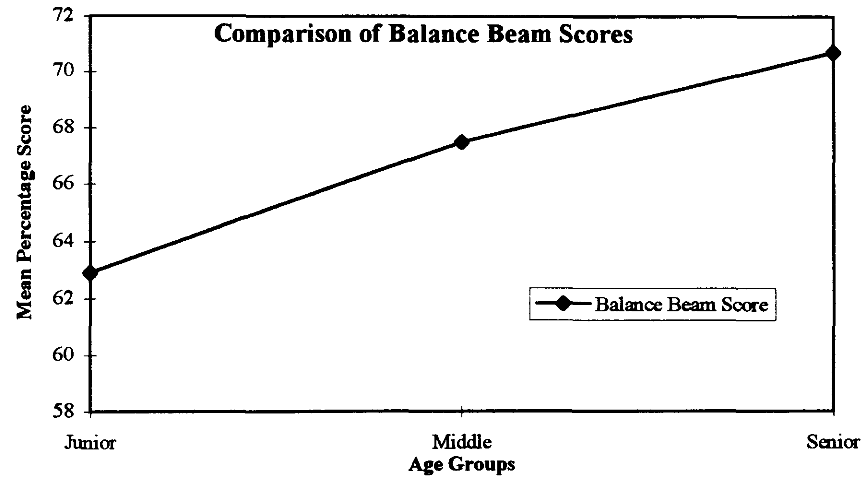


Figure 19: Comparison of Balance Beam Scores Across Three Age Groups

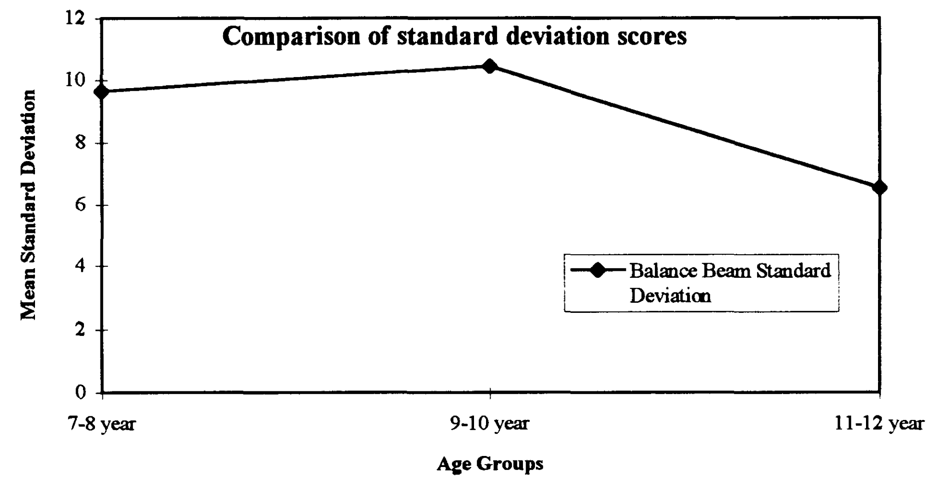


Figure 20: Comparison of Standard Deviations of Balance Scores

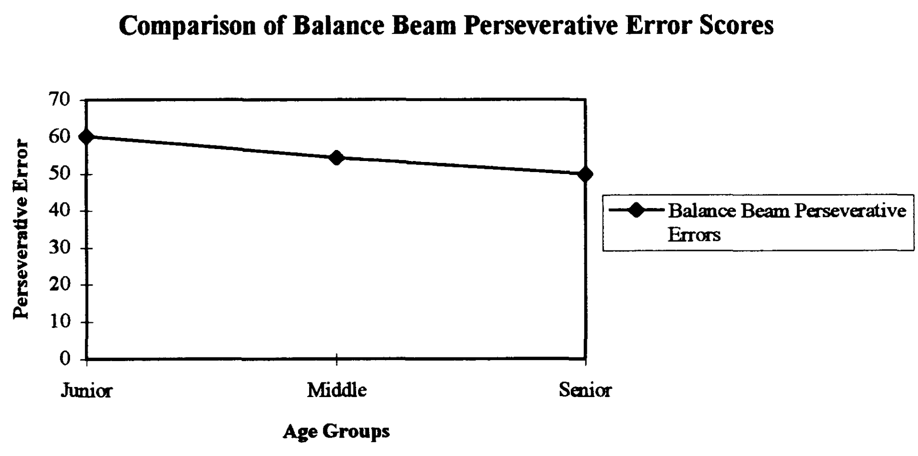


Figure 21: Comparison of Balance Beam Perseverative Error Scores

Item difficulty levels using Rasch analysis were developed across all subjects. For a full listing of item-difficulty Rasch scores and internal reliability coefficients, see Appendix 14. An analysis of individual trials using the Rasch difficulty scores indicates a two-stage level of difficulty with approximately half the trials in each stage. The first twelve trials consistently scored within the "too orderly" category with Rasch difficulty scores greater than 2 standard deviations below the mean. An inspection of the corresponding mean scores across all subjects confirms this. With the exception of Trial 9 which recorded a mean score of 0.885, all remaining trials in this bracket were >0.920. The mean score for this group of twelve trials was 0.960.

Two points arise from this result:

- the first eight trials were all of the Dominant type with solution based on number of disks. It has previously been established that the Dominant problem types occurred when unequal disks were placed on corresponding pegs from the fulcrum. The solution is based on one dimension - the number of disks on each side of the fulcrum.

- the next four trials were of the Subordinate type. Subordinate problem types involved equal disks which were placed on non-corresponding pegs. Solution is also one dimension- distance from the fulcrum. In these four trials, disks were equal on both sides of the fulcrum although placed on noncorresponding pegs.

If subjects continued to respond by making judgments based on the number of disks, as required for the first eight trials, then on Trial 9 they would have judged the result to be horizontal, which was incorrect. This may account for the low score on Trial 9 where an unexpected shift in criterion occurred.

The second difficulty stage involves the remaining twelve trials with the exception of Trial 18. Rasch difficulty scores for this bracket of trials were within 2 standard deviations above and below the mean (range of 2.179 to -1.709) suggesting that these eleven items were discriminating well and were a good fit to the Rasch unidimensional model. An inspection of mean scores reveals a range between 0.126 and 0.820 with a mean for this bracket of 0.422. Trials in this second half of the Balance Beam are of the Conflict-Subordinate and Conflict-Equal type with solution based on taking account of both number of disks and distance from the fulcrum.

Trial 18, which represents a third level of difficulty, is the first trial involving unequal disks placed on noncorresponding pegs from the fulcrum where the result is horizontally balanced. Rasch difficulty score for this trial was 5.404, suggesting that this trial does not conform to the other trials. Mean score across all subjects was 0.005 indicating that very few subjects were correct for this trial.

A Cronbach alpha internal reliability coefficient across all 183 subjects was computed using Systat. The coefficient was 0.561. The alpha coefficient result may have been depressed because of the effects of the two difficulty levels within the test. To partition this effect, internal reliability measures were computed for each of the two sections of the Balance Beam across the 183 subjects. Cronbach alpha reliability coefficient for the first difficulty level, Trials 1-12, was 0.635. The reliability coefficient for the second difficulty level, Trials 13-24, was 0.445. The difference in results suggests that the first half of the Balance Beam was responded to in a more consistent manner than the second half of the task. There are clearly at least two difficulty levels in performance on the Balance Beam task.

5.5 SEALS

Three measures were produced from the raw scores: Percent Correct, Seals Monitoring Score, and Seals Error Score. The Percent Correct was derived by adding the scores obtained for each of the three problem types and converting this to a percentage. The maximum score for each problem type, based on successful performance on the first two trials, was 6, which gave a possible maximum score of 18 for the three problem types.

For the Seals Monitoring score, all anecdotal information during testing was recorded and then coded according to the criteria detailed in Table 21. The Seals Monitoring Score was a composite of three graded responses which, on face value, were indicative of monitoring behaviours: querying the rules, restating the rules, and self-corrections. The rules and the scoring system are outlined in Table 21 below. The three types of monitoring activity are identified and represent increasing levels of monitoring behaviours. For statistical purposes, the three levels have been treated as mutually exclusive categories. The lowest level, (aside from the case when a subject is totally unaware or forgets a rule) is when a subject queries a rule. The subject has indicated an uncertainty of the actual rule and seeks confirmation. The next level of competency is when a subject is aware of the rule and restates it as a means of using that information to direct behaviour. This is consistent with Luria's (1966) concept of verbal regulation of behaviour.

The third level of monitoring behaviour is self-correction, which is the application of rule awareness in making adjustments or corrections to responses which have been tested to be inconsistent with the stated rules. From the anecdotal information used to classify the different types of monitoring activity, it is evident that those subjects who used 'querying the rules' as the only form of monitoring, displayed more uncertainty in their responses than those subjects who used the other two levels of monitoring. From the evidence of subjects' verbalisations, an arbitrary decision was made to devise a scoring system to reflect the levels of monitoring activity. For each of the three problem types (2-ball, 3-ball and 4-ball), a score was assigned to each of the three levels. This score is outlined in Table 21 .

Table 21 : Scoring system for Seals Monitoring Activity

- querying a rule was scored 1 point
- restating a rule was scored 3 points
- self-corrections were scored 5 points.

For each problem type, the maximum score was 5 and the minimum 0. Total scores over the three problem types therefore ranged from 0 to a maximum score of 15. The weightings given to each of the three monitoring activities were based on the notion that they represented different levels of cognitive processing. Results were converted to a percentage. The Seals Error score was based on the percentage of trials with rule violations.

The Seals Error score was defined by the percentage of trials with rule violations, taken over the three separate problems. Rule violations occurred when subjects moved more than one ball at a time, moved a ball from the bottom of a stack or placed a larger ball on top of a smaller ball. Once one of these violations occurred, then that particular trial was counted as an error trial.

Table 22: Mean scores for Seals Task

Age Group	Seals (% correct)	Seals Monitoring (%)	Seals Error (%)
Junior	48.3 (13.1)	15.355 (16.390)	55.435 (28.431)
Middle	60 (11.65)	11.7 (12.385)	51.383 (31.012)
Senior	68.2 (13.25)	25.623(18.305)	42.246(34.749)

Independent samples t-tests using Bonferroni adjustments are detailed below:

Table 23: T-tests for Seals Correct Score

Age Group	T-Score	D.F.	p-value
Junior-Middle	-4.946	120	0.000**
Junior-Senior	-7.906	121	0.000**
Middle-Senior	-3.499	119	0.001**

** significant at the Bonferroni-adjusted confidence level.

The results of Tables 22 and 23 indicate significant developmental improvement in scores across the three separate age groups. Seals Correct Score, based on the total score of the three problem types, appears to discriminate significantly across time and suggests that this task is a significant developmental measure for the three particular age groups. The Bonferroni adjusted values are significant at the $p < 0.017$ confidence level. An inspection of the standard deviations from Table 22, and from Figure 25, indicates that for Seals Correct performance, the 9-10 year old group displayed less variability in their performance than the other two age groups. One would expect a decrease in variability with age, along with an increase in performance. It is apparent that there is not a smooth linear relationship in degree of variability. It is proposed to take this point up in the Discussion. The standard deviations for the three age groups were 13.1, 11.65 and 13.25 for the Junior, Middle and Senior age groups, respectively.

Seals monitoring statistic is sensitive to developmental changes between the 9-10 year age group and the 11-12 year age group. There is no significant improvement in scores between the 7-8 and 9-10 year age group. There is also a significant difference in scores between the 7-8 year group and the older 11-12 year group and between the 9-10 year group and the 11-12 year group. This result suggests that monitoring behaviour during Seals performance displays a developmental improvement up to 9-10 years of age and then from 9-10 to 11-12 years of age. An inspection of Figure 22 highlights the decrement in performance of the 9-10 year old group suggesting that this group of subjects was somehow performing quite differently to the other age groups.

The Seals Error score indicates a *post hoc* improvement with age in rule violations, although this effect was not significant at the Bonferroni level. Figure 24 depicts the downward gradient with increasing age for Seals Error, providing support for the notion that breaking of rules occurs less often with increasing age.

An internal reliability score was not computed for Seals as the task did not conform to the format for a Cronbach Alpha reliability statistic. Not all subjects completed the same number of trials as criterion for completion was two successive successful trials for each problem type. A test-retest paradigm to obtain a measure of reliability was considered, although the possible effects of practice would more than likely contaminate the novelty value of the instrument.

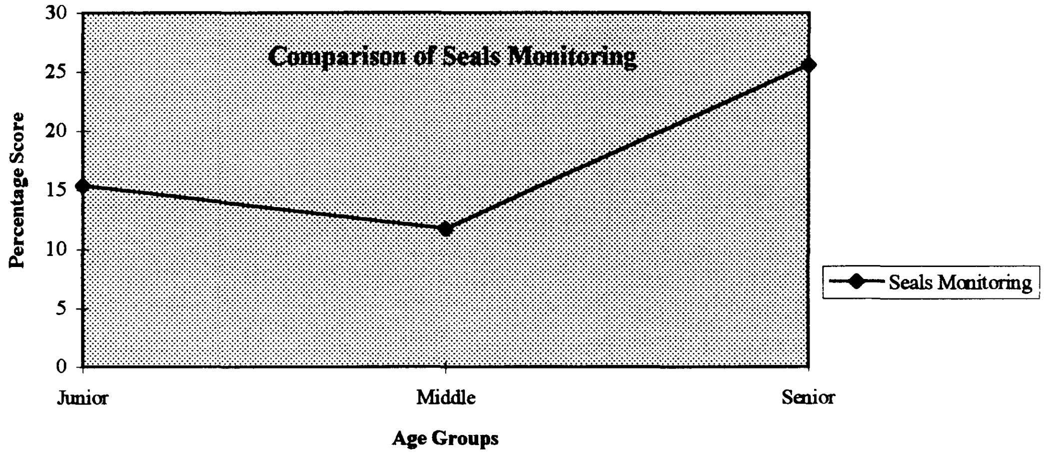


Figure 22: Comparison of Seals Monitoring scores across three age groups

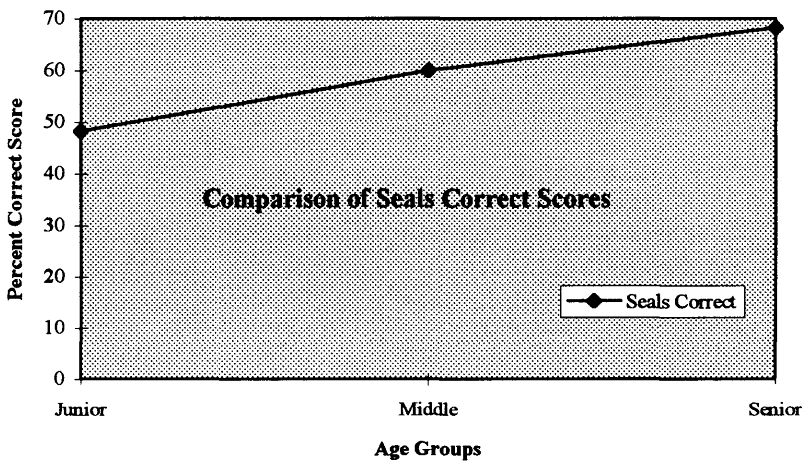


Figure 23: Comparison of Seals Correct Scores across three age groups

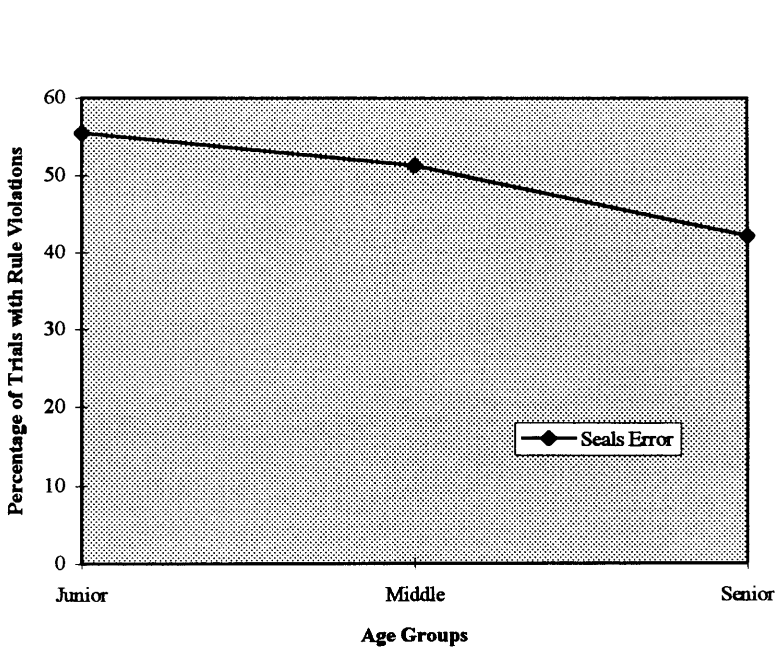


Figure 24: Comparison of Seals Error Scores across three age groups

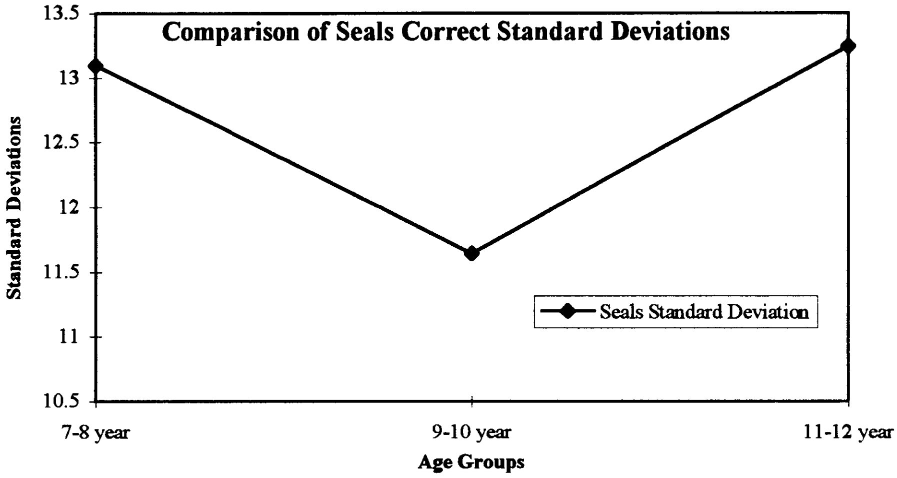


Figure 25: Comparison of Seals Standard Deviation scores across three age groups

Table 24: Seals Monitoring Score

Age Group	T-Score	D.F.	p-value
Junior-Middle	1.386	120	0.168
Junior-Senior	-3.279	121	0.001**
Middle-Senior	-4.892	119	0.000**

** significant at Bonferroni-adjusted confidence level

Table 25: Seals Error Score

Age Group	T-Score	D.F	p-value
Junior-Middle	0.753	120	0.453
Junior-Senior	2.289	120	0.024
Middle-Senior	1.525	119	0.130

Error-Types and Problem-solving Strategies

Individual differences for each of the 3 and 4-ball problem types were examined.

Table 26: Number of subjects classified according to score received as a function of problem type

	3-BALL						4-BALL					
Groups	6	5	4	3	2	0	6	5	4	3	2	0
7-8	5	14	8	16	8	11	0	0	0	0	1	59
9-10	16	22	12	7	1	2	0	2	1	1	0	56
11-12	22	24	8	6	1	0	2	6	2	5	10	36

Table 26 details the number of subjects in each score category for the 3 and 4-ball problem types. Scores are based on the trials at which criterion (two in a row) was reached. For example, a score of 6 was given for passing the first two trials in the minimum number of moves, a score of 5 for passing the second and third trials, and so on. As the number of subjects within each age group was comparable (62, 60, 61), a direct comparison was made for each score for each problem type. It is evident that all the 11-12 year olds were able to pass the 3-ball problem, whereas eleven of the 7-8 year olds were unsuccessful. Only two of the 9-10 year olds were unsuccessful on the 3-ball problem. The 4-ball problem posed a greater challenge for all three age groups. Few subjects scored 6, or even 5 on the 4-ball problem, and while twenty-five of the 11-12 year old subjects were able to solve the problem within six trials, only one 7-8 year old and four 9-10 year olds were able to do so.

The key errors for the 3 and 4-ball problem tasks were identified by examining the errors both within and across the age groups. The pattern of errors was taken from the first error which occurred in each of the two trials of the 3-and 4-ball problems. This gave an opportunity to examine the range of errors in Trial 2 when the first move was given to the subjects, and ensured that all subjects' results would be counted as ceiling was reached on two successive successful attempts.

For both the 3 and 4-ball problems, the error pattern was consistent across the age groups with the first move posing the greatest difficulty. Although twice as many 12 year olds incurred no errors on Trial 1, their overall error pattern was the same as the 7 year olds. The critical key error points, across both problems and across all age groups occurred at Moves #1, #3, #5, and #9. It appeared that despite the obvious higher success rates by the older age groups, their critical choice points occurred at the same moves as the younger, less sophisticated, problem-solvers. This point is illustrated clearly in Tables 29 and 30. During performance of the 4-ball problem, the majority of errors occurred on Move #1 during Trial #1, with between 52 and 69 percent of subjects incurring errors on this move. On Trial #2, between 29 and 46 percent of subjects incurred errors on Move #5. Tables 29 and 30 also reveal that a higher percentage of 11-12 year old subjects incurred errors on these critical choice points in comparison to the two younger age groups. This point will be taken up in the Discussion. The percentage of errors for each move for each age group is detailed in the tables below, with the trials listed separately.

Table 27: Percentage of Subjects Exhibiting Trial 1 Errors for 3-ball Seals.

	7-8yrs	9-10yrs	11-12yrs
Move			
#1	74	72	57
#2	3	0	0
#3	10	3	5
#4	0	0	3
#5	2	0	0
Error free	11	25	35

Table 28: Percentage of Subjects Exhibiting Trial 2 Errors for 3-ball Seals

Move	7-8yrs	9-10yrs	11-12yrs
#2	9	3	6
#3	42	24	10
#4	3	1	2
#5	10	10	5
Error free	36	62	77

Table 29: Percentage of Subjects Exhibiting Trial 1 Errors for 4-ball Seals

Move	7-8yrs	9-10yrs	11-12yrs
#1	52	54	69
#2	10	8	0
#3	24	8	8
#4	1	3	1
#5	13	22	7
#7	0	1	0
#9	0	4	8
Error free	0	0	7
n	30	31	30

Table 30: Percentage of Subjects Exhibiting Trial 2 Errors for 4-ball Seals

	7-8yrs	9-10yrs	11-12yrs
Move			
#2	27	6	7
#3	31	25	13
#4	2	0	0
#5	29	44	46
#6	1	0	1
#7	3	0	7
#9	7	19	9
#10	0	1	0
Error Free	0	5	17

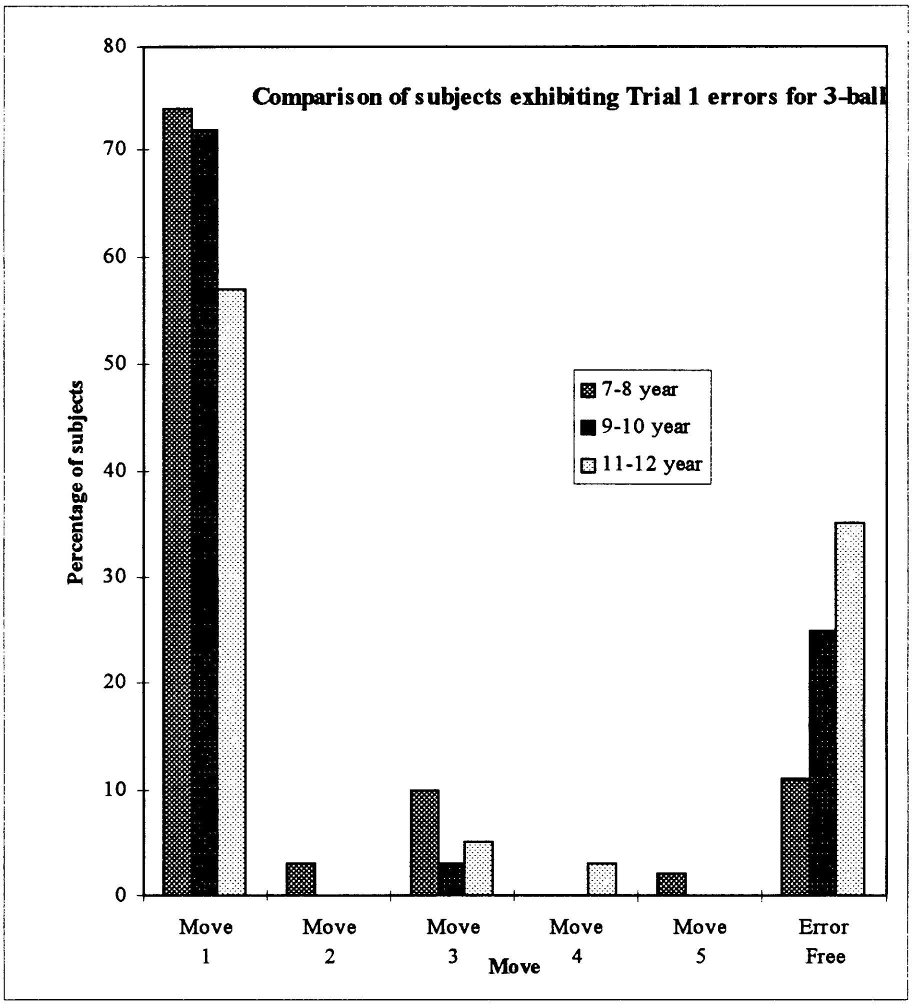


Figure 26: Comparison of percentage of subjects exhibiting Trial 1 errors on the 3-ball Seals problem

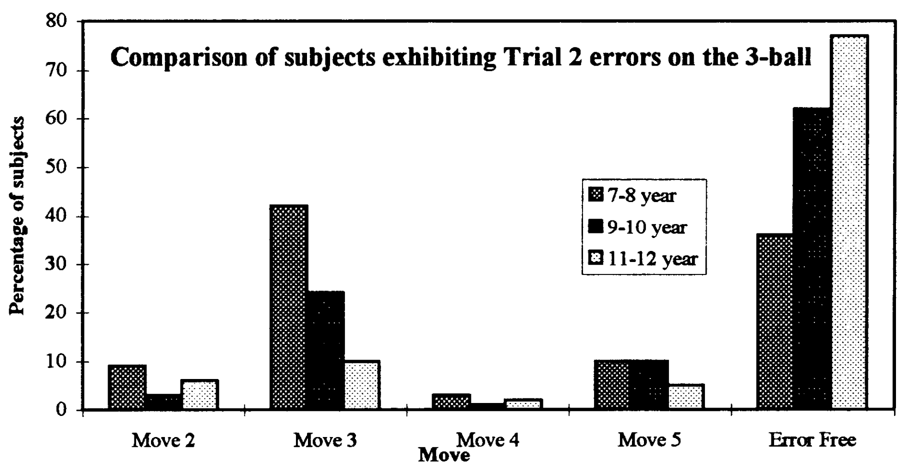


Figure 27: Comparison of percentage of subjects exhibiting Trial 2 errors on Seals 3-ball problem

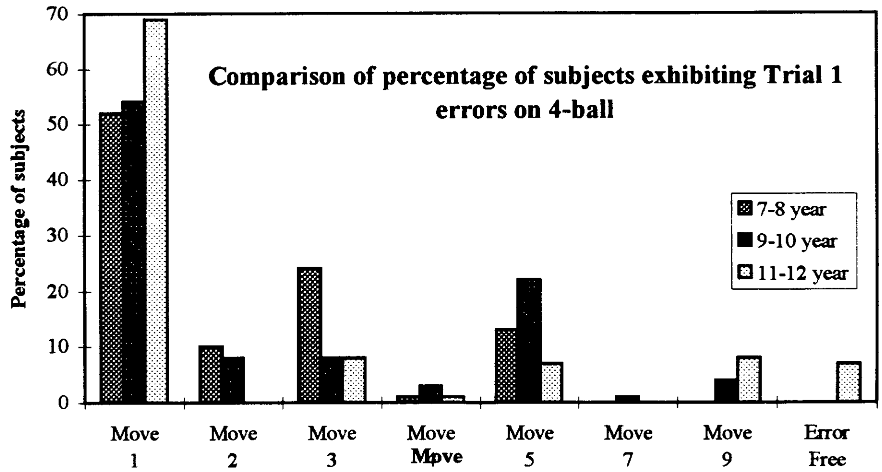


Figure 28: Comparison of percentage of subjects exhibiting Trial 1 errors on Seals 4-ball problem

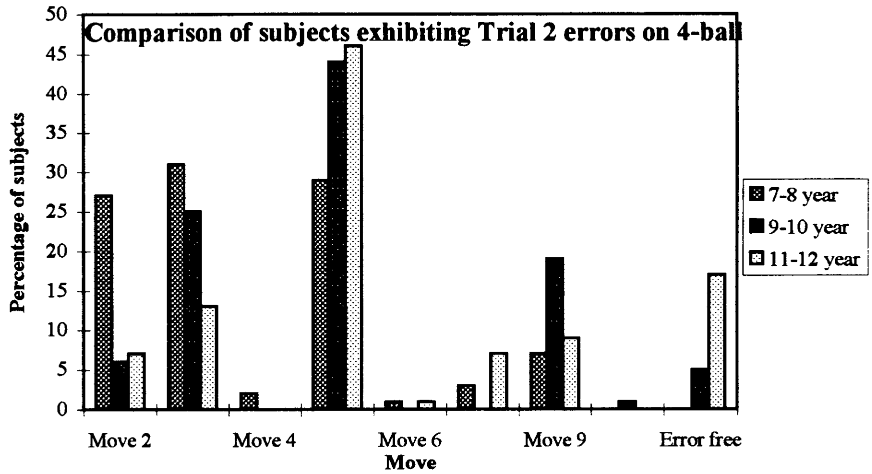


Figure 29: Comparison of percentage of subjects exhibiting Trial 2 errors on Seals 4-ball problem

3-Ball

It is apparent, from Tables 27 and 28, and Figure 26, that proportionate to those subjects who made errors, the percentage of those who made first move errors was comparable for each age group for the 3-ball problem. When the first move was given to subjects, in Trial 2, Move #3 became a key error point. This occurred across all the age groups. For those who mastered Move #3, Move #5 was the next critical point in the problem.

4-Ball

The key error patterns on Trials 1 and 2 of the 4-ball problems in Tables 29 and 30 and graphed in Figures 28 and 29, present a similar picture. The significant key error, reflected in a higher error rate, occurred on Move #1, with Moves #3, #5, and #9 representing critical decision points in the problem. As you move from the youngest age group to the oldest across the tables, it is apparent that the 7 and 8 year olds had great difficulty with the task with a significant proportion of their errors occurring on the first move of both trials (27%). None of the subjects at this age group successfully solved the task. Compare this with the 11-12 year old subjects, 40% of whom successfully solved the problem within the six trials. Despite the relative success of the older age group, their key error points remained at Moves #1, #3, #5, and #9.

Self-corrections

An analysis of the self-corrections made by all subjects during the first two trials of the 3 and 4-ball problems is enlightening. Only the first self-correction for each subject was included in the analysis and, as self-corrections appeared to be evenly distributed across both problem types, the results were aggregated. For each age group, self-corrections appeared only at the key error points of Moves #1, #3, and #9. The incidence of self-corrections appeared stable for the 7-8 and 9-10 year age groups, and increased for the 11 and 12 year olds.

It was also at these same moves where the majority of verbalisation occurred. A verbalisation was defined as any utterance that occurred during performance of a particular move. It ranged from a single word or sound to a string of sentences. All of these utterances during a single move were classified as one instance of verbalisation. All utterances were recorded verbatim for classification. For example, on Move #3 during Trial 1 of the 3-ball problem, Subject 218 asked: "Can you move the balls backward?" After thinking about the solution for a while, he moved correctly. Subject 215, on Trial 2 of the 3-ball problem said: "Big one can't go on a small one - that's a problem! I don't know what to do with the Blue one!" For this subject, this was the only point in this trial where verbalisation occurred. The numbers of self-corrections for all age groups, with the 3 and 4-ball problems combined are outlined in Table 31:

Table 31: Incidence of self-corrections for both 3 and 4-ball Seals

Age	Move #1	Move #3	Move #9
7-8 years	8	1	0
9-10 years	7	2	0
11-12 years	13	2	4

Self-corrections only occurred on the moves outlined in Table 31 above. It is evident that, as a group, the older age group made more self-corrections than the other two groups, and that for all three groups, significantly more self-corrections occurred at Move #1. It has previously been established that Move #1 posed the greatest difficulty for all age groups and for both problem types. Table 31 confirms that the older age group was using more self-monitoring strategies, particularly on the most difficult choice point in the problem resolution.

5.6 WISCONSIN CARD SORTING TASK (WCST)

Four scores were developed from the results of the WCST: Number of correct sorting sets (Correct), Perseverative Errors, Total Correct responses, and Time taken to complete the task. Mean scores and standard deviations for each of the four scores are listed in Table 32. Results of *t*-tests comparing performance on the WCST across the three age groups indicate that, statistically, there was a lack of incremental developmental progression across each of the three groups. There are developmental differences, but these are not demonstrated statistically across the three age groups. For example, at the Bonferroni adjusted level, WCST Time taken displayed a developmental improvement from 9-10 years of age. From Figure 34, it is evident that there is a steady decrease with increasing age in the time taken to complete the WCST. The performance of the younger age group did not differ significantly to that of the 9-10 year group. Taken together with the results detailed in Table 35, it is suggested that although there is no significant difference across the three age groups in the number of correct sets achieved, as graphed in Figure 23, subjects did improve in their speed of responding from the age of 9-10 years of age. The older age group was quicker in completing the WCST task, although did no better. In fact, a *post hoc* comparison of mean correct set scores from Table 32, and an inspection of Figure 31, indicates that the Middle age group had the highest mean score of the three age groups (3.43 correct sets, compared to 3.26 for Senior and 2.90 for Junior).

An inspection of the level of variability in performance, detailed in Table 32 and Figure 35, indicates that for WCST error scores, the 9-10 year old group had standard deviations which were larger than expected. For example, on Perseverative Error, the 9-10 year group had the largest standard deviation score (5.24, 5.82 and 4.95 for Junior, Middle and Senior groups, respectively). For the Total Errors on the WCST, the 9-10 year group's standard deviations were not dissimilar to that of the 7-8 year group (8.9 and 8.75 for Junior and Middle groups). In addition, a *post hoc* inspection of the number of Correct Sets performance indicates an increasing variability with age, with an increase in standard deviation of performance with increasing age This result was unexpected.

It appeared that the 9-10 year old group was somehow functioning quite differently to the two other age groups in terms of their variability in performance on the WCST. There was a trend toward improvement on WCST Perseverative Errors between the Junior and Senior Groups although this was not significant at the Bonferroni adjusted level of confidence.

Table 32: Scores on Wisconsin Card Sorting Test

Age Group	Correct	Perseverative Error	Total Error	Time
Junior	2.90 (1.08)	9.27 (5.24)	21.47 (8.9)	374.25 (122.77)
Middle	3.43 (1.23)	7.00 (5.82)	17.83 (8.75)	349.88 (83.53)
Senior	3.26 (1.46)	7.27 (4.95)	17.84 (9.75)	276.76 (59.77)

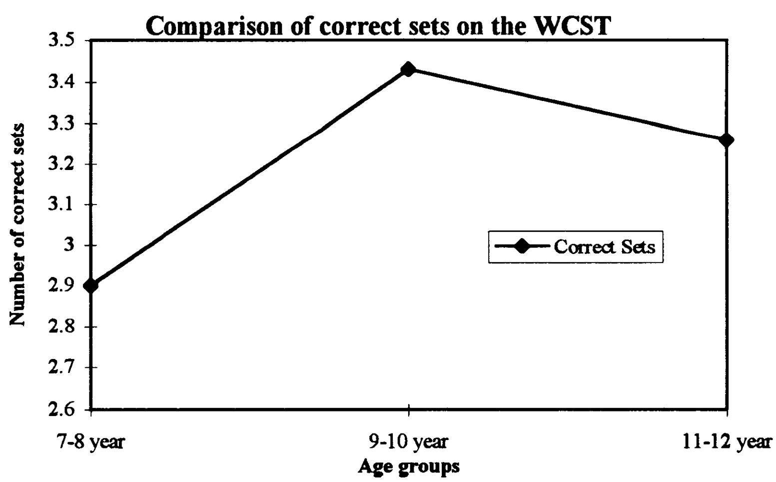


Figure 31: Comparison of WCST correct sets scores across three age groups

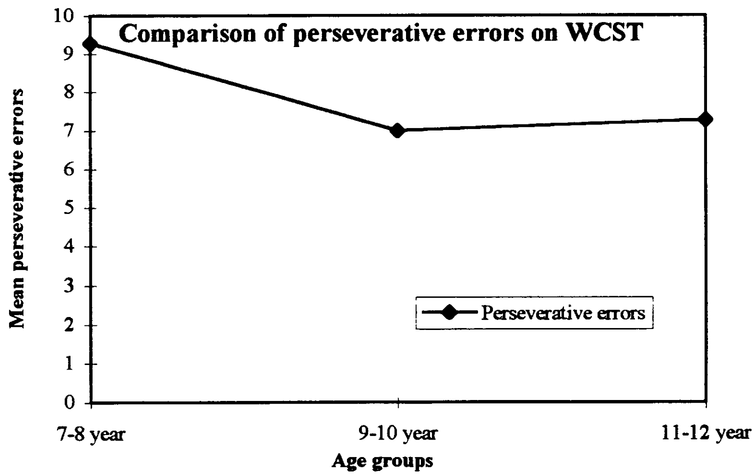


Figure 32: Comparison of WCST Perseverative Error scores across three age groups

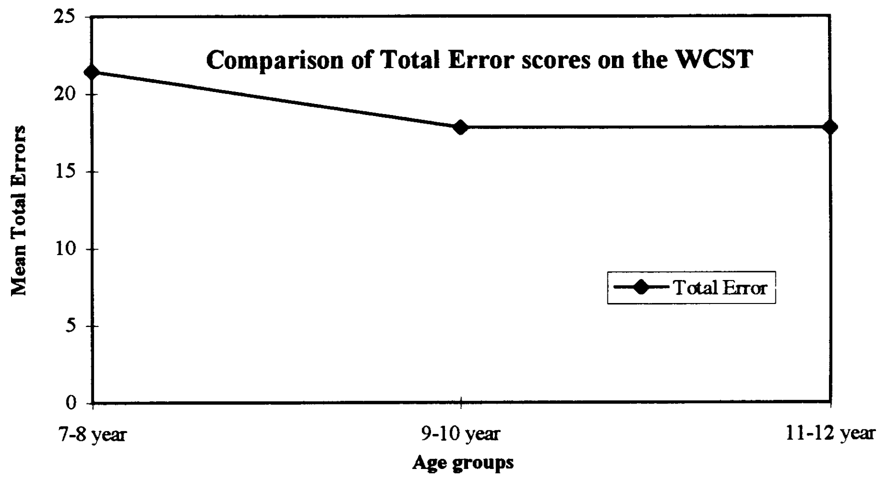


Figure 33: Comparison of WCST Total Error Scores across three age groups

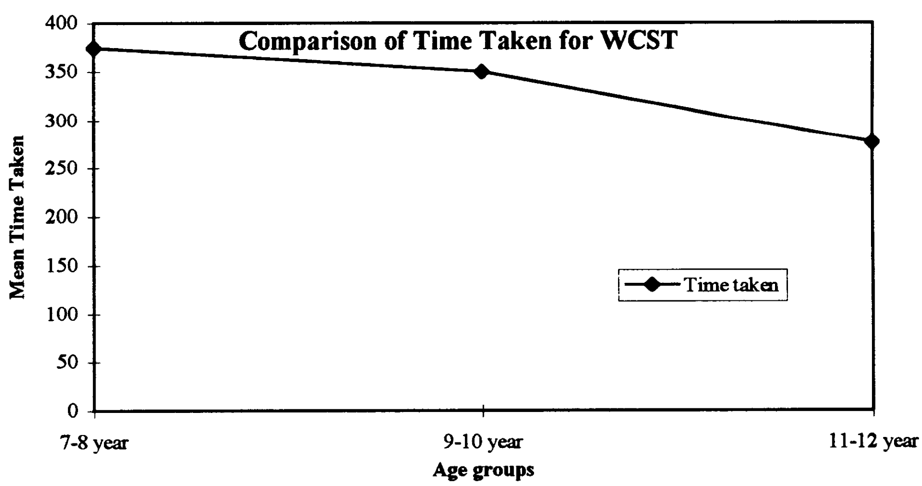


Figure 34: Comparison of WCST Time Taken scores across three age groups

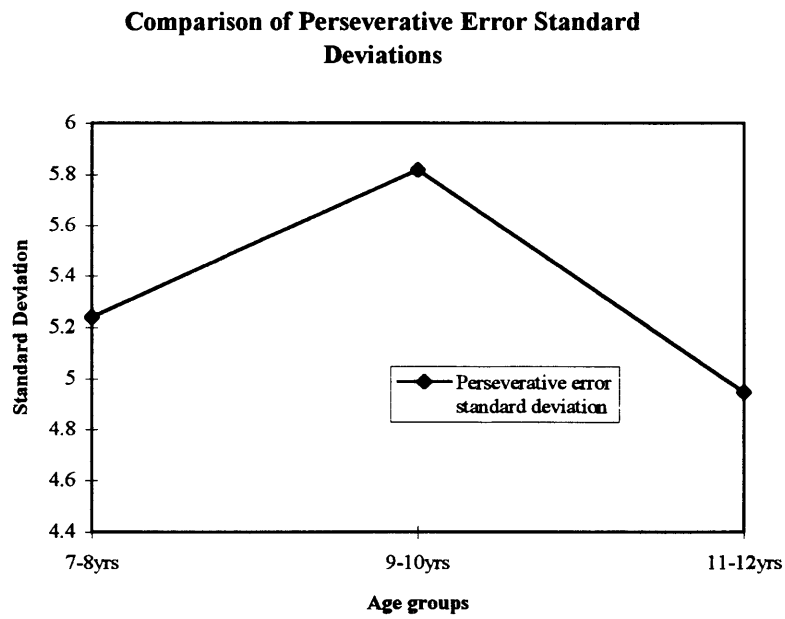


Figure 35: Comparison of WCST Perseverative Error standard deviation scores across three age groups

Independent sample t-tests using Bonferroni adjustments are detailed in Table 33.

Table 33: T-tests for WCST Correct Score

Age Groups	T-Score	D.F.	p-value
Junior-Middle	-1.916	116	0.058
Junior-Senior	-1.783	119	0.077
Middle-Senior	0.006	117	0.995

Table 34: T-tests for WCST Perseverative Errors Score

Age Groups	T-Score	D.F.	p-value
Junior-Middle	2.214	116	0.029
Junior-Senior	2.161	120	0.033
Middle-Senior	-0.165	117	0.870

Table 35: T-tests for WCST Time Taken

Age Groups	T-Score	D.F.	p-value
Junior-Middle	1.503	116	0.136
Junior-Senior	5.604	120	0.000**
Middle-Senior	5.026	117	0.000**

** Significant at Bonferroni-adjusted confidence level

5.7 AUDITORY VERBAL LEARNING TEST (AVLT)

A summary of the means and standard deviations (brackets) for the results of the AVLT are detailed in Tables 36 and 37. Figure 36 details the individual learning rates for the three age groups over the first five trials of the AVLT. It is evident that the three curves from Figure 36 describe classical learning curves as mean numbers of words recalled increase for each of the three age groups, with each succeeding presentation trial. The performance of the 7-8 year and 9-10 year old subjects follows a similar rate of recall with their learning curves quite close to each other. The older 11-12 year old subjects describe a smoother curve with their scores clearly separated from the other two groups. The Acquisition Rate and Learning Scores, described in Figures 37 and 38, serve to emphasise the significant difference in cumulative and total words remembered over the five trials between the 11-12 year old subjects and the other two age groups. In both Figures the graphs describe a relatively flat profile between the 7-8 and 9-10 year groups, and then rises sharply to the 11-12 year old group, suggesting a significant increase in performance for the older group. Independent t-tests, outlined in Table 40, statistically confirm the significant difference at the $p<0.017$ confidence level between the Junior and Senior age groups.

Acquisition Rate refers to the rate of increase in the number of words recalled over successive trials during Trials 1 to 5. Proactive Interference refers to the tendency for previous learning to interfere with recall of new information. This is measured by comparing the number of words recalled by the subject in Trial 1 and the number of words recalled by the subject in Trial 6 and expressing these as a ratio. Retroactive Interference refers to the tendency of new learning to interfere with the recall of previously learned material. A measure of retroactive interference is given by the ratio of recall on Trial 7 with recall on the last learning trial (Trial 5).

Retrieval Efficiency refers to the capacity to retrieve words from memory storage. Retrieval efficiency is measured as the ratio of the proportion of Trial 8 words recalled, to the signal detection measure of recognition performance (p(A)) (Geffen, Sjobert, Mason, Smyth and Butterworth, 1993). According to Geffen *et al* (1993), p(A)= 0.5 (1 + HR-FP), where HR is the proportion of list words that are correctly recognised, and FP (False Positive) refers to an incorrect positive response to a "new" word in the recognition test.

Table 36: Mean scores and standard deviations for measures of AVLT

Age Group	Acquisition	Proactive	Retroactive	Forgetting	Retrieval
Junior	1.23 (0.48)	1.07 (0.44)	0.81 (0.27)	0.97 (0.33)	0.55 (0.21)
Middle	1.24 (0.50)	1.01 (0.42)	0.82 (0.25)	0.99 (0.26)	0.57 (0.19)
Senior	1.41 (0.53)	1.04 (0.44)	0.83 (0.16)	0.99 (0.23)	0.67 (0.19)

Mean scores and standard deviations (in brackets) for Learning (total correct words recalled for lists 1-5), Recognition A (total words correctly recognised from List A), and Recognition B (total words correctly recognised from List B) are detailed in Table 37.

Table 37: Mean scores and standard deviations for scores on the AVLT

Age Group	Learning	Recognition A	Recognition B
Junior	37.629 (9.9)	12.855 (2.6)	6.161 (2.7)
Middle	39.586 (8.9)	13.466 (1.5)	6.534 (22.3)
Senior	44.295 (7.9)	13.787 (1.7)	6.689 (2.8)

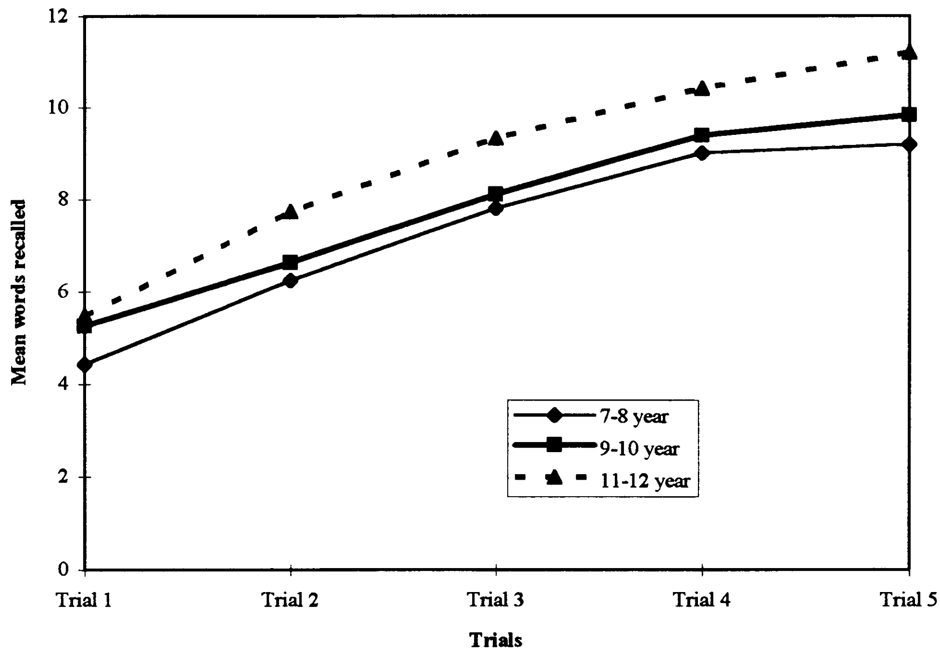


Figure 36: Comparison of words recalled on each of five trials of the AVLT across three age groups

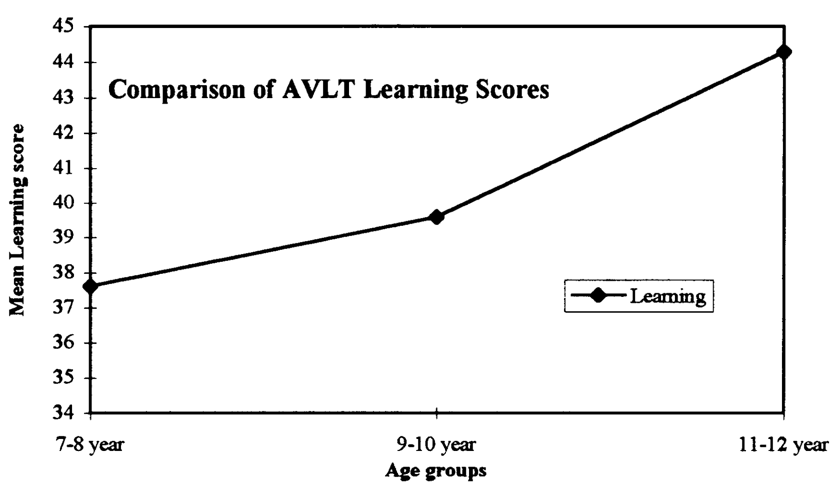


Figure 37: Comparison of AVLT words recalled over five trials (Learning Score) across three age groups

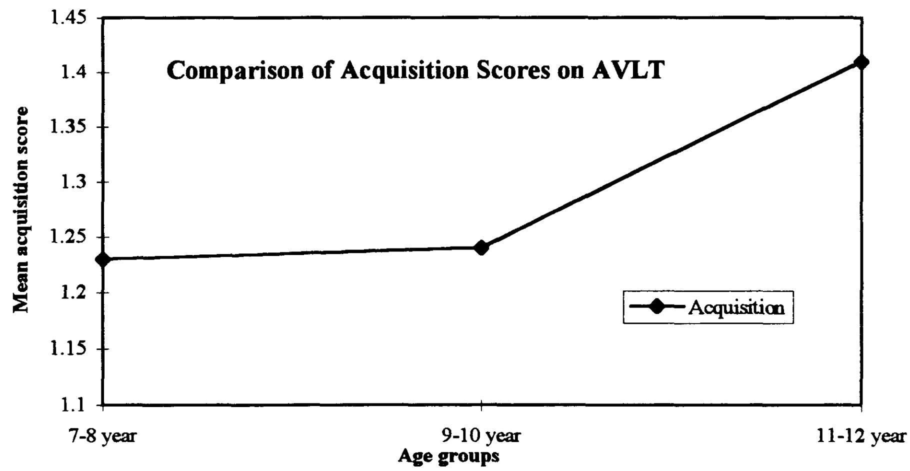


Figure 38: Comparison of AVLT Acquisition scores across three age groups

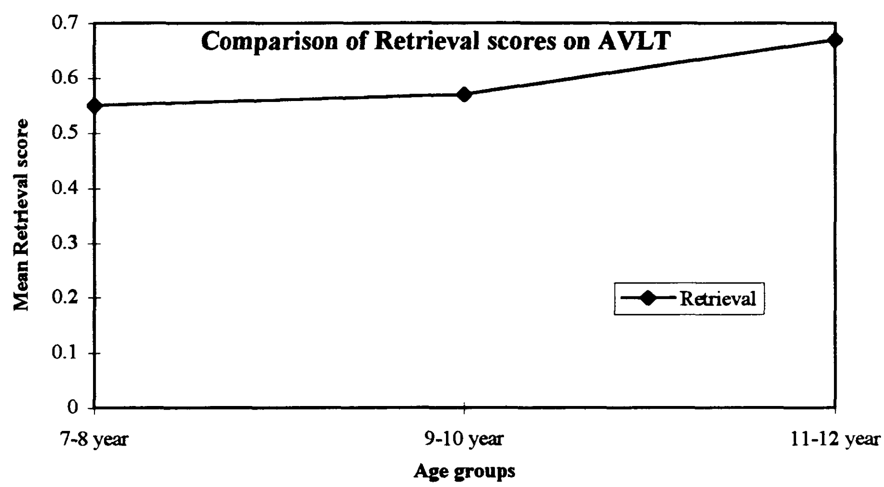


Figure 39: Comparison of AVLT Retrieval scores across three age groups

Paired comparison t-tests involving independent samples with Bonferroni adjustments are displayed in Tables 38-45:

Table 38: Paired comparison of ACQUISITION Scores

Group	T-Score	D.F.	p-value
Junior-Middle	-0.091	119	0.928
Junior-Senior	-2.012	121	0.046
Middle-Senior	-1.852	118	0.067

Table 39: Paired comparisons for AVLB

Group	T-Score	D.F.	p-value
Junior-Middle	-0.815	116.3	0.417
Junior-Senior	-1.062	121	0.290
Middle-Senior	-0.333	114.5	0.740

Table 40: Paired comparisons for LEARNING

Group	T-Score	D.F.	p-value
Junior-Middle	-1.134	118	0.259
Junior-Senior	-4.120	121	0.000**
Middle-Senior	-3.041	113.7	0.003**

** significant at Bonferroni-adjusted confidence level

Table 41: Paired Comparisons for RECOGNITION

Group	T-Score	D.F.	p-value
Junior-Middle	-1.571	118	0.119
Junior-Senior	-2.381	121	0.019
Middle-Senior	-1.111	116.7	0.269

Table 42: Paired comparisons for FORGETTING

Group	T-Score	D.F.	p-value
Junior-Middle	-0.381	119	0.704
Junior-Senior	-0.464	121	0.644
Middle-Senior	-0.068	118	0.946

Table 43: Paired comparisons for PROACTIVE INTERFERENCE

Group	T-Score	D.F.	p-value
Junior-Middle	0.855	120	0.394
Junior-Senior	0.428	121	0.670
Middle-Senior	-0.414	119	0.680

Table 44: Paired comparisons for RETROACTIVE INTERFERENCE

Group	T-Score	D.F.	p-value
Junior-Middle	-0.121	119	0.904
Junior-Senior	-0.552	121	0.582
Middle-Senior	-0.429	118	0.668

Table 45: Paired comparisons for RETRIEVAL

Group	T-Score	D.F.	p-value
Junior-Middle	-0.471	119	0.638
Junior-Senior	-3.206	121	0.002**
Middle-Senior	-2.816	118	0.006**

** significant at Bonferroni-adjusted confidence level.

Results of the independent t-tests for the various measures of the AVLT, indicate that two measures, LEARNING of five repeated trials of words and RETRIEVAL efficiency of recalling words following a delay period, display developmental improvement. In both cases, significantly improved performance occurs with age from the 9-10 year group to the 11-12 year group. The performance of the younger 7-8 age group is not significantly different from that of the 9-10 year group although is significantly different from that of the senior group. The results suggest that for both subtests of the AVLT, performance significantly improves from 9-10 years of age. A *post hoc* inspection of mean scores across the three age groups indicates a trend toward improvement with age. The remaining six scores on the AVLT did not discriminate across the three age groups. These results are consistent with the findings by Forrester and Geffen (1991) who demonstrated a significant difference between Learning and Retrieval indices between 7-8 years and the two older 9-10 and 11-12 year olds, which both performed at a comparable level. In addition, no significant difference is found on the other indices. It appears that the number of words remembered over five trials and retrieval efficiency are both sensitive to developmental improvement with age in a population of 7-12 year old healthy children.

There is, however, a consistent difference in scores between the present study and those of Geffen and her colleagues. The 9-10 year old subjects in the current analysis scored one standard deviation below Geffen's subjects in both Learning and Retrieval indices on the AVLT. Both samples came from Australian groups which were representative of the Australian population, both involved male subjects in mainstream classes and both groups were administered the same version of the AVLT according to Lezak (1983). However, the Geffen sample only contained 10 subjects, compared to the 61 in the current analysis. The possibility for a sample-biased difference is significant. The two results are outlined in Table 46:

Table 46: Comparison of AVLT mean scores with Forrester and Geffen's (1991) results

	Learning		**Retrieval**	
	Bogan	Geffen	Bogan	Geffen
7-8 Years	37.6 (9.9)	39.1 (5.9)	0.55 (0.21)	0.56 (0.15)
9-10 Years	39.6 (8.9)	47.3 (5.8)	0.57 (0.19)	0.71 (0.13)
11-12 Years	44.3 (7.9)	47.0 (4.8)	0.67 (0.19)	0.73 (0.09)

5.8 AUSTIN MAZE

Nine scores were generated from the Austin Maze test. These were: Total Errors, Time, Mean Errors/Trial, and five error types. The results of these scores are detailed in Tables 47 and 48:

Table 47: Means and standard deviations of measures of the Austin Maze.

Age Group	Total Errors	Time	Mean Errors/Trial
Junior	135.21 (67.88)	1064.15 (408.52)	8.77 (7.88)
Middle	107.395 (56.68)	851.36 (311.615)	6.05 (2.38)
Senior	53.715 (47.52)	567.18 (183.04)	4.66 (1.90)

Table 48: Means and standard deviations of error types from the Austin Maze.

Age Group	Error Type 1	Error Type 2	Error Type 3	Error Type 4	Error Type 5
Junior	10.97 (7.0)	4.10 (1.65)	42.57 (9.16)	2.58 (2.27)	4.78 (3.25)
Middle	9.89 (7.47)	3.80 (1.78)	40.85 (8.36)	2.71 (2.13)	3.67 (2.85)
Senior	7.67 (7.48)	3.67 (1.98)	37.72 (10.79)	3.33 (2.43)	2.83 (1.88)

Independent sample T-Tests using Bonferroni adjustments were conducted and the results are detailed in Tables 49 to 55.

Table 49: T-Tests for Austin Maze Total Error Score

Age Groups	T-Score	D.F.	p-value
Junior-Middle	2.363	118	0.020
Junior-Senior	6.369	120	0.000**
Middle-Senior	4.377	116	0.000**

** significant at Bonferroni-adjusted level.

Table 50: T-Tests for Austin Maze Error Type 1

Age Groups	T-Score	D.F.	p-value
Junior-Middle	0.794	112	0.429
Junior-Senior	2.453	114	0.016**
Middle-Senior	1.575	110	0.118

** significant at Bonferroni-adjusted confidence level.

Table 51: T-Tests for Austin Maze Error Type 2

Age Groups	T-Score	D.F.	p-value
Junior-Middle	0.940	112	0.349
Junior-Senior	1.290	114	0.200
Middle-Senior	0.375	110	0.708

Table 52: T-Tests for Austin Maze Error Type 3

Age Groups	T-Score	D.F.	p-value
Junior-Middle	1.076	119	0.284
Junior-Senior	2.679	120	0.008**
Middle-Senior	-1.767	117	0.080

** significant at Bonferroni-adjusted confidence level

Table 53: T-Tests for Austin Maze Error Type 4

Age Groups	T-Score	D.F.	p-value
Junior-Middle	-0.322	112	0.748
Junior-Senior	-1.735	114	0.085
Middle-Senior	-1.443	110	0.152

Table 54: T-Tests for Austin Maze Error Type 5

Age Groups	T-Score	D.F.	p-value
Junior-Middle	1.929	112	0.056
Junior-Senior	3.948	114	0.000**
Middle-Senior	1.866	110	0.065

** significant at Bonferroni-adjusted confidence level

Recall that the error types are as follows:

Error Type 1:	Does not go back to the last correct square
Error Type 2:	Back tracking
Error Type 3:	Perseverative error with respect to the last trial
Error Type 4:	Diagonal moves
Error Type 5:	Perseverative error within trial.

It is clear from the results of Tables 50-54, that three of the five error types display some sensitivity to developmental changes across the three age groups. There are significant differences between the 7-8 year olds and the 11-12 year olds on three error measures: does not go back to the last correct square (Error Type 1), perseveration with respect to the last trial (Error Type 3), and perseveration within the same trial (Error Type 5). In addition, a *post hoc* inspection of mean scores for these three error measures, depicted in Table 48, indicates an improvement in scores (less errors) with increasing age. Perseveration, both within and across trials, appears to be age-related, with a decrease in perseverative responses with increasing age.

The three age groups did not differ in rule-breaking behaviours, as measured by Error Type 2 and Error Type 4. An inspection of Table 48 reveals a somewhat unexpected result. Incidence of illegal diagonal moves actually increases with increasing age across the three age groups, although this result is not significant. This result may reflect a failure to inhibit.

Table 55: T-Tests for Austin Maze Mean Errors/Trial

Age Groups	T-Score	D.F.	p-value
Junior-Middle	2.107	118	0.037
Junior-Senior	3.272	120	0.001**
Middle-Senior	3.499	116	0.001**

** significant at Bonferroni-adjusted confidence level

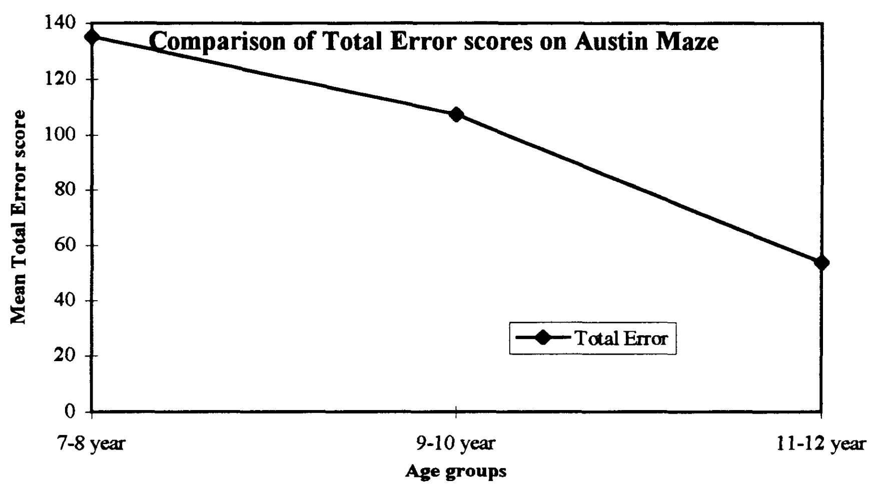


Figure 40: Comparison of Austin Maze Total Error scores across three age groups.

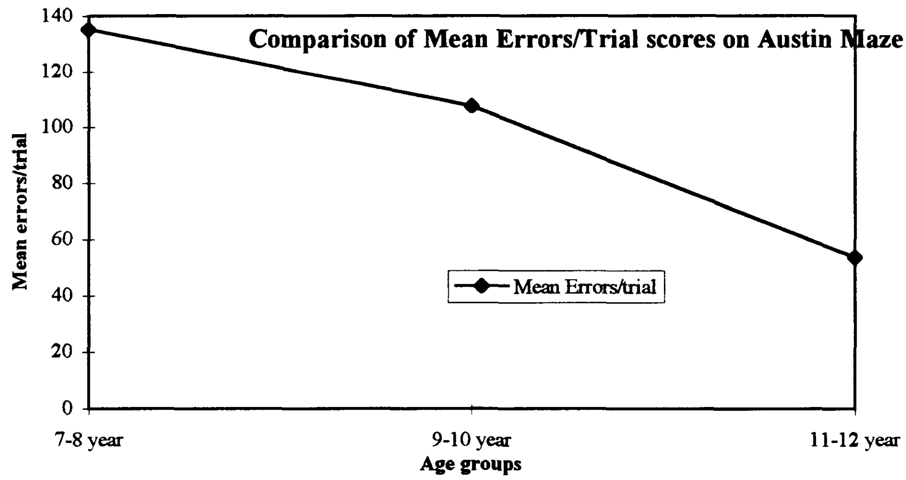


Figure 41: Comparison of Austin Maze Mean Errors/Trial scores across three age groups.

The results of Austin Maze error scores indicate that three measures have some degree of significance. Mean errors per trial is a score which significantly discriminates between the 9-10 year age group and the 11-12 year age group. A *post hoc* comparison of the mean scores reveals that the Senior group made significantly less errors per trial than the Middle age group. There is also a significant difference between the Junior and the Senior groups. Figure 41 depicts the developmental trend in performance for Mean Errors/Trial. The Maze Total Error scores also indicate a similar pattern of development in that there is a significant difference in scores between the Middle and Senior groups of students and between the Junior and Senior groups. Figure 40 graphs this result.

Trials to Criterion

A direct comparison of the number of trials administered to each age group (maximum of 20 trials) indicates no difference in performance between the 7-8 and 9-10 year olds. The 11-12 year age group, on the other hand, required less trials to be administered in order to reach criterion. Mean trials administered were 17.9, 17.1 and 13.9 for the 7-8, 9-10 and 11-12 year age groups, respectively. The percentage of subjects, within each age group who did in fact reach criterion within the 20 trials highlights the superior performance of the 11-12 year old subjects: 38.3%, 40.3% and 80% of 7-8, 9-10 and 11-12 year olds reached criterion within the given trials. Figure 42 below highlights the difference in performance across the three age groups.

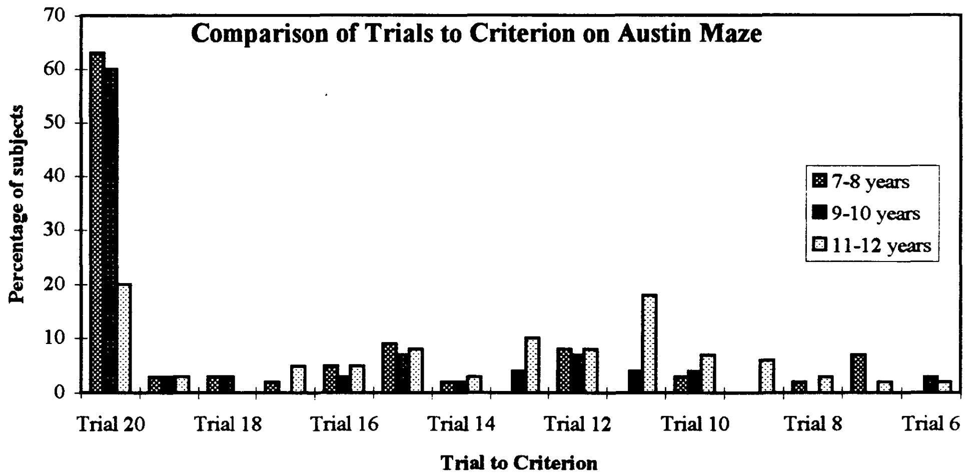


Figure 42: Percentage of subjects who reached criterion across number of trials on the Austin Maze

Error Utilisation

"Error utilisation" (Walsh, 1985) is based on the notion that subjects with frontal impairment take many more trials to reach two consecutive error-free trials than normal subjects. Subjects with prefrontal lobe or executive function deficits would fail to monitor their performance and to use environmental cues to adjust their behaviour. As a consequence, on the Austin Maze, it is hypothesised that these subjects fail to take their errors into account and to rapidly improve their performance. A comparison was made of subjects' error scores on the first ten trials with those of the total errors over the twenty trials or when criterion was reached. To what extent were subjects' scores on the first ten trials predictive of later success?

Pearson correlation coefficients were computed for each age group and the results indicate that for the Middle and Senior age groups the correlations are 0.955 and 0.956 respectively. For the Junior group, however, the correlation is $r= 0.775$. These results indicate that for those subjects at least 9 years of age, scores from the first ten trials are a good estimate of scores to criterion. However, the picture is not quite so clear for the younger age group. It cannot be predicted with such certainty that scores for the 7-8 year old subjects over the first ten trials are a good estimate of overall scores to criterion. It appears that the performance profile of the 7-8 year old group is qualitatively different from the two older age groups of subjects on the Austin Maze.

5.9 OLSAT

The OLSAT test has been normed on a mean score of 100 and a standard deviation of 16 (Otis and Lennon, 1989). It is apparent, from Figure 43, that the Senior Age group is closest to the normative sample on the three scales of the OLSAT. According to the manual, the Senior age group scored in the average range of school ability on the Verbal, Nonverbal and Total scales. Both the Junior and Middle Age groups are within one standard deviation of the normative data from the OLSAT manual. Both groups fall within the low average range of school ability on all three scales. It is considered that the three age levels did not differ significantly from a normal population of children in terms of school ability. However, the 7-8 and 9-10 year old groups, particularly, are a little on the low side. It is noted that the demographic data, detailed in Table 9, indicate that the Normative Sample is a little underrepresented in the areas of parents with tertiary education and those holding down a white collar job. For example, the Sydney City Sample has 13.4% with a tertiary degree as the highest educational qualification, compared with 6.7% of the Normative sample with a tertiary degree. In addition, 32% of the Sydney City Sample was holding down a white collar job compared to 20.6% of the Normative Sample. These two factors may have introduced some sample bias. In terms of standard deviations, it appears that the 9-10 year old group is closest to the normative sample listed in the manual, with a standard deviation of 15.82 for the Total Score, compared with 16 for the normative sample.

Means and standard deviations for each age group for the Total Score, Verbal Score and Nonverbal Score are detailed in Table 56:

Table 56: Results of OLSAT

Age Groups	Total Score	Verbal	Nonverbal
Junior	88.24 (14.9)	86.28 (13.63)	89.66 (14.96)
Middle	88.25 (15.82)	87.76 (14.75)	88.00 (13.92)
Senior	97.19 (13.34)	94.66 (12.15)	93.48 (13.35)

6.1.1 Memory / Learning Factor

Tests which have significant loadings on the Memory/Learning factor are the Verbal and NonVerbal OLSAT school ability tests, and the AVLT Learning variable. It is contended that these variables rely on memory skills for competent performance. An analysis of the skills involved in the two OLSAT scales indicates that successful performance relies on memory of school learning. The AVLT Learning score is derived from the number of words correctly remembered over five learning trials. This is a measure of immediate memory for words.

6.1.2 Executive Function

Factor 2, Executive Function, comprises Balance Beam, Seals and WCST Correct, which load positively and Austin Maze time taken and Austin Maze mean errors per trial which load negatively on this factor. The loadings on the two Austin Maze variables are consistent with the two developmental tests of Balance and Seals and the WCST Correct in that low scores are indicators of successful performance. An analysis of the skills involved in tests which load on the Executive Function factor reveals a common thread involving planning, monitoring, mental flexibility and active memory. Both Seals and Balance Beam, on *a priori* grounds, were specifically developed to test the interaction of active memory and planning skills. The Austin Maze is purported to be a measure of planning, monitoring and error utilisation (Walsh, 1987). Successful performance on the WCST requires mental flexibility and the ability to use environmental feedback to adjust behaviour (Lezak, 1983).

6.1.3 Inhibition

Factor 3, Inhibition, taps into Fish and Piggy Bank both of which have significant loadings on this factor. Both tests are hypothesised to have a significant motor set/inhibition component in terms of inhibiting previously correct and inappropriate responses. Only those variables which loaded significantly on the separate factors were then selected according to the magnitude of their loadings on the factor. Only those variables with a numerical value of 0.5 or greater were considered. The following run was then executed, based on the factor loadings from Table 60.

Let Factor 1 = Verbal + Nonverbal + 0.5 AVLT - 0.5 WCSTime

Let Factor 2 = Balance + Seals + WCST Correct - AUSmeane - AUStime

Let Factor 3 = Fish + Piggy Banks.

The Factor equations are based on the magnitude of the loadings. Factor 1 equation, for example, is derived from the fact that both Verbal and Nonverbal variables have loadings in the order of 0.9, whereas AVLT and WCST Time are closer to 0.5. To adjust for this discrepancy, 0.5 was multiplied by the AVLT and WCST Time scores. The variable loadings on the remaining two factors did not display this level of difference and a correction is not considered necessary. In the strictest sense, the resultant "factors" are actually linear combinations of variables suggested by the factor loadings. However, with this caveat in mind, they will continue to be referred to as 'factors'. Scores on each factor are standardised on the basis of total variation with the result that for each factor, the overall mean is zero and the overall standard deviation is one.

The MANOVA output indicates that for each of the three factors, the mean factor scores increase with increasing age. For the factors labelled Executive Function and Inhibition, the standard deviations decrease with age suggesting less variability in scores with increasing age. Means and standard deviations are listed in Table 61:

Table 61: Means and Standard Deviations for Factor Scores for Three Age Groups

Junior Group	**Memory**	**Executive Function**	**Inhibition**
Mean	-0.201	-0.647	-0.386
SD	0.990	0.906	1.241
Middle Group			
Mean	-0.158	0.145	0.072
SD	1.071	0.733	0.831
Senior Group			
Mean	0.376	0.631	0.373
SD	0.880	0.695	0.638

Pair-wise comparisons were then computed for the three factors across the three age groups. A Bonferroni adjustment of alpha/9 (three groups x three factors) was set to establish a stringent upper limit of confidence for Type 1 errors. The results indicate a significant difference in factor scores at the $p<0.006$ level (Bonferroni adjusted), across each of the three groups for the Executive Function factor. Given the stringent Bonferroni adjustment, it can be concluded that the Executive Function factor unequivocally maximises the separation between the age groups. The Executive Function factor is the only one which displays a significant developmental improvement across the three age groups.

The Memory/Learning factor has a significant difference in factor scores between the Junior age group and the Senior age group, and between the Middle and Senior groups, although not for the Junior and Middle groups. Performance on tests which loaded onto the Memory/Learning factor demonstrate a significant improvement from the age of 9-10 years of age. There is no significant difference in performance between 7-8 and 9-10 years of age on tests which load on the Memory/Learning factor. The Response Inhibition factor displays a significant difference in scores only between the younger and older age groups. Subject's scores on tests which load onto the Inhibition factor did not significantly differ between 7-8 and 9-10 years of age or between 9-10 and 11-12 years. Inhibition is consequently not significantly tied to a clear developmental timetable of improvement in test scores. Table 62 contains the output for the F-tests.

Table 62: Tests for Effects across Factor Scores for each of the Three Age Groups

Age Group	Factor	DF	F	P
	Memory/Learning	1	0.052	0.820
Junior-Middle	Executive	1	27.273	0.000*
	Inhibition	1	6.318	0.013
	Memory/Learning	1	10.080	0.002*
Junior-Senior	Executive	1	76.675	0.000*
	Inhibition	1	18.798	0.000*
	Memory/Learning	1	7.991	0.005*
Middle-Senior	Executive	1	10.262	0.002*
	Inhibition	1	2.750	0.099

* significant at the $p<0.006$ confidence level with Bonferroni adjustment.

The MANOVA canonical loadings (correlations between dependent variables and the discriminant function that maximises the differences between groups) over the entire 183 subjects were inspected from the output. The Executive Function loading is 0.922 This is a highly significant correlation and indicates that the Executive Function factor unambiguously accounts for the majority of variance of the discriminant function that separates the groups over the entire sample of children from 7 years to 12 years. Canonical loadings are reproduced below:

Memory/Learning	0.303
Executive Function	0.922
Inhibition	0.455

Pearson correlation analyses were then conducted on the three factors for each of the three age groups.

The results of these factor correlational matrices indicate an increasing correlation of factors with age. There is clearly a greater independence of factors at the younger age group. The largest correlations are between Memory/Learning and Executive Functions, and the strength of this correlation increases from 0.309 to 0.457 to 0.596 with increasing age. The relationship between Executive Functions and Response Inhibition also increases relatively. These correlations are: 0.019, 0.186, and 0.319 for the 7-8 year, 9-10 year and 11-12 year age groups respectively. The correlations between the Memory/Learning factor and Inhibition factor are not as stable, nor significant. The results for each age group are detailed in Table 63:

Table 63: Pearson Correlation Matrices for three factors within each age group

Junior Group (7-8years)

	Memory Learning	Executive Function	Inhibition
Memory /Learning	1.000		
Executive Function	.309	1.000	
Inhibition	0.123	-0.019	1.000

Middle Group (9-10years)

	Memory Learning	Executive Function	Inhibition
Memory	1.000		
Executive Function	0.457	1.000	
Inhibition	0.034	0.186	1.000

Senior Group (11-12years)

	Memory Learning	Executive Function	Inhibition
Memory	1.000		
Executive Function	0.596	1.000	
Inhibition	0.221	0.319	1.000

A Fisher Transformation R to Z analysis (Hays, 1981) was then calculated to test for significant differences between the correlations across the three age groups. The significance of the correlations between Executive Function and Memory/Learning for the Junior and Senior age groups ($r=0.309$ and $r=0.596$) produced a two-tailed value of $p=0.048$, which is significant at the $p<0.05$ level.