# 7.0 RELATIONSHIP BETWEEN SCHOOL ABILITY, MEMORY, AND EXECUTIVE FUNCTION

## 7.1 Executive Function and School Ability

From the underlying Factor Structure it has been demonstrated that Seals, Balance Beam and to a lesser extent, WCST Correct, load positively and significantly onto an Executive Function factor, with loadings of 0.504, 0.526 and 0.452, respectively. In addition, Austin Maze Mean Errors/Trial and Austin Maze Time Taken load negatively onto the Executive Function factor with loadings of – 0.728 and -0.610 respectively. The four tests, taken together, have significant loadings on Executive Function for the age groups sampled. To test the relationship of school ability performance, as measured by the OLSAT Total Score, to executive function, a Pearson correlation coefficient was generated for each of the tests which load significantly on the Executive Function factor, and the OLSAT Total Score for each age group. The outcomes, in addition to the correlations of all the developmental tests to OLSAT scores, are outlined in Table 64.

**Table 64: Correlations between OLSAT Total score and other tests for three Age Groups**

| Tests | Junior | Middle | Senior |
|---|---|---|---|
| Ausmean | 0.140 | -0.306 | -0.427 |
| Austime | -0.344 | -0.549 | -0.620 |
| AVLTLearn | 0.530 | 0.310 | 0.507 |
| Balance | 0.355 | 0.403 | 0.622 |
| Fish | 0.061 | 0.140 | 0.333 |
| Piggy | 0.099 | 0.182 | 0.055 |
| Seals | 0.383 | 0.325 | 0.359 |
| WCSTCorr | 0.005 | 0.359 | 0.250 |
| WCSTtime | -0.095 | -0.470 | -0.210 |
| OLSAT | 1.000 | 1.000 | 1.000 |

The first observation to be made from the above correlation tables is that there appears to be an increasing and positive correlation, across the three age groups, between OLSAT Total scores and ability on the Balance Beam Task. The correlations for the three age groups are r = 0.355, r = 0.434, and r = 0.641 for the Junior, Middle and Senior age groups respectively. There is also an increasing negative correlation across the three groups for both Austin Maze variables, which translates into a positive relationship between performance on the Austin Maze and the school ability measure, as both Maze measures are error scores. The relationship between Seals and the OLSAT Total score remains stable across the three age groups with correlations of $r = 0.383$, $r = 0.325$ and $r = 0.359$ for the 7-8, 9-10 and 11-12 year age groups respectively.

The correlation between OLSAT Total Score and the AVLT Learning variable (the total words recalled over the five learning trials), not surprisingly has a positive score with correlations of $r = 0.530$, $r = 0.310$ and $r = 0.507$ for the three age groups respectively. Both OLSAT and AVLT load onto the Memory/Learning factor.

For the three age groups, correlations between OLSAT Total and the remaining tests listed in Table 64 are not significant, indicating that high or low scores on OLSAT Total are not associated with high or low scores on these tests. For example, the correlation between performance on the OLSAT Total and the Piggy Bank task approaches zero indicating that the two tests are quite independent of one another. The Piggy Bank task taps into delayed response and inhibition, concepts which play very little part in performance on the OLSAT test.

To test the statistical significance of the correlation coefficients across the three age groups among the Balance Beam task, the two Austin Maze variables and OLSAT, a Fisher Transformation R to Z was computed (Hays, 1981). The results indicate a significant difference, at the $p<0.05$ confidence level, between the correlations involving the Junior and Senior age groups for the scores on the Balance Beam task and Austin Maze time taken and those on the OLSAT Total variable. In addition, there is a significant difference in correlations between the Junior and Senior and Junior and Middle age groups for Austin Maze mean errors per trial and the OLSAT scores. These results indicate that, with increasing age, from 7-8 years to 11-12 years, performance on the Balance Beam task increasingly depends upon school ability. At the 11-12 year age group, for example, there is a significant correlation between performance on the Balance Beam task and performance on the OLSAT Total, with $r = 0.641$. In addition, time taken to complete the Austin Maze test becomes increasingly related to success on the OLSAT test, with a reduction in time with increasing age. Tables 65 and 66 detail these outcomes.

**Table 65: Significance of correlations between Balance Beam and OLSAT Total across the three age groups.**

| Group | r | p (2-tailed test) |
|---|---|---|
| Junior-Middle | 0.355-0.434 | 0.614 |
| Middle Senior | 0.434-0.641 | 0.114 |
| Junior-Senior | 0.355-0.641 | 0.036* |

* significant at p<0.05 level

**Table 66: Significance of correlations between Austin Maze variables and OLSAT Total across the three age groups**

|  | Ausmean Errors |  | Austime |  |
|---|---|---|---|---|
| Group | r | p | r | p |
| Junior-Middle | 0.14 -0.306 | 0.01* | -0.344 -0.549 | 0.166 |
| Middle-Senior | -0.306 -0.427 | 0.454 | -0.549 -0.620 | 0.562 |
| Junior-Senior | 0.14-0.427 | 0.001* | -0.344 -0.620 | 0.048* |

* significant at p<0.01 level

To assess the predictive validity of estimating OLSAT scores from the executive measures, regression analyses investigating both linear and nonlinear relationships were computed using Systat. The nonlinear model was restricted to a quadratic relationship. Within each age group the regression analyses were generated separately for Seals and Balance, with OLSAT Total Score as the dependent variable.

Results indicate a significant linear relationship between Balance and OLSAT and Seals and OLSAT for all three age groups. The nonlinear models for both sets of relationships across the three age groups are not significant indicating that they do not provide any predictive validity over the linear models. There is therefore no evidence of curvature in the regression analyses indicating that the linear relationship is sufficient to account for the data. Tables 67 and 68 summarise the regression outcomes.

**Table 67: Regression analyses predicting OLSAT from Balance scores within three age groups.**

| Groups | Model | T | P (2-tail) |
|---|---|---|---|
| Junior | Linear | 2.839 | 0.006* |
| | Quadratic | 1.437 | 0.156 |
| Middle | Linear | 3.218 | 0.002* |
| | Quadratic | 1.081 | 0.285 |
| Senior | Linear | 6.308 | 0.000* |
| | Quadratic | -0.838 | 0.407 |

* significant at $p<0.05$ confidence level

**Table 68: Regression analyses predicting OLSAT scores from Seals scores for three age groups.**

| Group | Model | T | P (2-tail) |
|---|---|---|---|
| Junior | linear | 3.102 | 0.003* |
| | quadratic | 0.378 | 0.707 |
| Middle | linear | 2.386 | 0.021* |
| | quadratic | -0.896 | 0.374 |
| Senior | linear | 3.107 | 0.003* |
| | quadratic | -0.010 | 0.992 |

* significant at $p<0.05$ confidence level

## 7.2 Executive Function and Memory

A significant relationship between the underlying Memory/Learning factor and the Executive Function factor has already been demonstrated. The correlations between the two factors for the three age groups were reported at $r=0.309$, $r=0.457$ and $r=0.596$ for the Junior, Middle and Senior age groups. A Fisher R to Z analysis produced a significant difference, at the $p<0.05$ confidence interval, between the correlations for the Junior and Senior age groups. It appears that developmentally, memory plays an increasingly significant role in executive function performance.

Does the positive and increasing relationship between memory and executive function also hold true for individual tests? To examine the relationship between memory performance as measured by the AVLT Learning index, and executive function, as measured by Seals and Balance Beam, a Pearson correlation analysis of the ten variables used in previous analyses was computed using Systat. The results indicate no clear developmental trend in their relationship with the AVLT Learning index. Correlation coefficients in the order of 0.27 and 0.39 were recorded by the Balance and Seals variables and the Memory index. There is, however, a significant correlation between OLSAT and AVLT learning scores across the three age groups. This is not an unexpected result as there is considerable overlap between memory performance and school ability performance. The coefficients are listed in Table 69.

**Table 69: Correlations between AVLT Learning Scores and other tests for the three age groups.**

| Tests | Junior | Middle | Senior |
|---|---|---|---|
| Ausmean | 0.049 | -0.206 | -0.202 |
| Austime | -0.385 | -0.438 | -0.409 |
| AVLTlearn | 1.000 | 1.000 | 1.000 |
| Balance | 0.373 | 0.278 | 0.295 |
| Fish | -0.032 | 0.071 | 0.331 |
| Piggy | 0.193 | 0.118 | 0.004 |
| Seals | 0.395 | 0.031 | 0.272 |
| WCSTcorr | 0.173 | 0.246 | 0.155 |
| WCSTtime | -0.182 | -0.258 | -0.179 |
| OLSAT | 0.530 | 0.310 | 0.507 |

To measure the predictive validity of estimating AVLT memory scores from Balance Beam and Seals scores, regression analyses investigating both linear and quadratic models were computed using Systat. Significant regression equations predicting memory scores from both Seals and Balance indices were found for the Junior and Senior age groups only. The Middle age group did not yield a positive relationship for memory scores for either Balance or Seals scores. These results are summarised in Tables 70 and 71.

**Table 70: Regression analyses predicting AVLT Learning scores from Balance score within three age groups.**

| Groups | Model | T | P (2-tail) |
|---|---|---|---|
| Junior | Linear | 3.440 | 0.001* |
| | Quadratic | 0.572 | 0.570 |
| Middle | Linear | 1.448 | 0.153 |
| | Quadratic | 1.351 | 0.182 |
| Senior | Linear | 2.233 | 0.029* |
| | Quadratic | -1.005 | 0.319 |

* significant at p<0.05 confidence level

**Table 71: Regression analyses predicting AVLT Learning scores from Seals scores for three age groups.**

| Group | Model | T | P (2-tail) |
|---|---|---|---|
| Junior | linear | 3.345 | 0.001* |
| | quadratic | 1.482 | 0.144 |
| Middle | linear | 0.564 | 0.575 |
| | quadratic | 1.060 | 0.294 |
| Senior | linear | 2.037 | 0.046* |
| | quadratic | 0.418 | 0.677 |

* significant at p<0.05 confidence level

## 8.0 ACQUIRED BRAIN-INJURED SAMPLE

The subjects for the brain-injured sample were taken from those children referred to Westmead Hospital, a large trauma referral centre in Sydney. Subjects, 9-10 years of age, who were enrolled in mainstream classes and who had suffered a brain injury as a result of a trauma were selected for inclusion in the sample. From a possible cohort of 33 cases, permission forms were returned for 21 children. The final sample size was 20. Details of these subjects relative to the normative 9-10 year old Middle group are detailed in Table 72:

**Table 72: Characteristics of Brain-Injured Sample in Relation to Normative Sample**

| Demographic Data | Brain-injured Sample | Normative Sample |
|---|---|---|
| n | 20 | 60 |
| Males | 14 | 60 |
| Females | 6 | 0 |
| OLSAT School Ability | 86.3 | 88.9 |
| Mean Age (months) | 119.6 | 119.32 |

The ratio of females to males was 1:2 in the brain-injured sample. This is consistent with demographic data (Kraus, Rock and Hemyari, 1990). Mean age for each sample is remarkably similar. The two groups are also closely matched in terms of school ability, as measured by the OLSAT Total Score. Statistically, there is no significant difference between the two groups on this variable (F=0.053, p=0.818). For a complete listing of raw scores across all the tests administered, see Appendix 15.

### 8.1 Information Gained from Medical Records

Medical records of 16 of the 20 subjects are listed in Appendix 16. The remaining four records had been subpoenaed because compensation claims were pending and the subjects' files were consequently unavailable. The medical records were used to verify the presence of neurological involvement as a result of insult to the brain. It is apparent that this is not an homogeneous group with wide variability in age at time of accident, loss of consciousness, degree of severity of brain injury as measured by the Glasgow Coma Scale, and presence of lesions as indicated by CT scans. Of the 16 subjects, all had experienced an external force to the head either as an MVA or a fall, and all 16 had been hospitalised as a consequence of their injuries. Ten subjects were admitted to hospital with witnessed loss of consciousness, eight had Glasgow Coma Scale scores less than 8, indicating a severe brain injury, and twelve had abnormal brain CT scans, only three of which reported frontal lesions. At the time of their accident, two subjects were younger than 5 years of age, twelve were between the ages of 5 and 8 and two were older than 8 years of age. Mean length of time since the brain injury was two years ten months, with a range from seven months to 6 years, 9 months. Despite their injuries, all twenty subjects were attending mainstream classes at the time of testing.

## 8.2 Procedure:

Nineteen of the twenty children were administered the same battery of assessments under the same conditions that were employed with the normative sample. These children lived in the Western Region of Sydney, in the same geographical area as the normative sample, and were tested at their home schools. The twentieth child, who lived on the Central Coast of NSW some 80 kilometres from Sydney, was tested at Westmead Hospital in an interviewing room. The order of presentation of tests was preserved for this twentieth child.

## 8.3 Results

### 8.3.1 Brain-Injured Sample Compared with Matched Normative Sample

The overall results of the brain-injured sample, in comparison with the matched normative group, suggest a larger variability in individual test scores and a *post hoc* comparison of poorer test performance. These are summarised, test by test, and detailed in Table 73.

**Table 73: Mean scores, standard deviations (in brackets) and tests of significance between acquired Brain-injured and Normative Sample of 9-10 year olds.**

| Tests | Brain-injured Sample | Normative Sample | F | p-Value |
|---|---|---|---|---|
| Seals | 46.9 (13.9) | 59.9 (11.8) | 16.817 | 0.000* |
| Balance Beam | 63.8 (10.6) | 67.5 (10.5) | 1.369 | 0.246 |
| Fish | 90.4 (13.4) | 91.8 (11.4) | 0.210 | 0.648 |
| Piggy Bank | 89.5 (11.0) | 92.2 (9.2) | 1.008 | 0.319 |
| Austin Maze mean Errors | 6.9 (3.8) | 6.0 (2.4) | 1.917 | 0.171 |
| WCST Correct Sets | 2.6 (1.2) | 3.4 (1.2) | 3.825 | 0.055 |
| AVLT Learning | 36.5 (8.7) | 39.6 (8.9) | 1.666 | 0.201 |
| Austin Maze Time | 936.6 (335.6) | 851.4 (311.6) | 1.392 | 0.242 |
| WCST Time | 354.3 (71.6) | 349.9 (83.5) | 0.018 | 0.895 |
| OLSAT Total Score | 86.3 (12.9) | 88.9 (16.3) | 0.053 | 0.818 |

* significant at Bonferroni adjusted confidence level

### 8.3.2 Seals

Results indicate that the brain-injured sample, as a whole, performed significantly lower than the matched normative sample on the Seals Task. The difference in mean scores is significant at the Bonferroni adjusted level (F=16.817, p=0.000). A comparison of performance on the 3 and 4-ball problems is detailed in Table 74 below:

**Table 74: Percentage of subjects classified according to score received as a function of problem length**

|  | 3-Ball Problem |  |  |  |  |  | 4-Ball Problem |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | 6 | 5 | 4 | 3 | 2 | 0 | 6 | 5 | 4 | 3 | 2 | 0 |
| Brain-injured | 0 | 25 | 25 | 20 | 5 | 25 | 0 | 0 | 0 | 0 | 0 | 100 |
| Normative | 27 | 37 | 20 | 11 | 2 | 3 | 0 | 3 | 2 | 2 | 0 | 93 |

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 of moves, a score of 5 for passing the second and third trials, and so on.

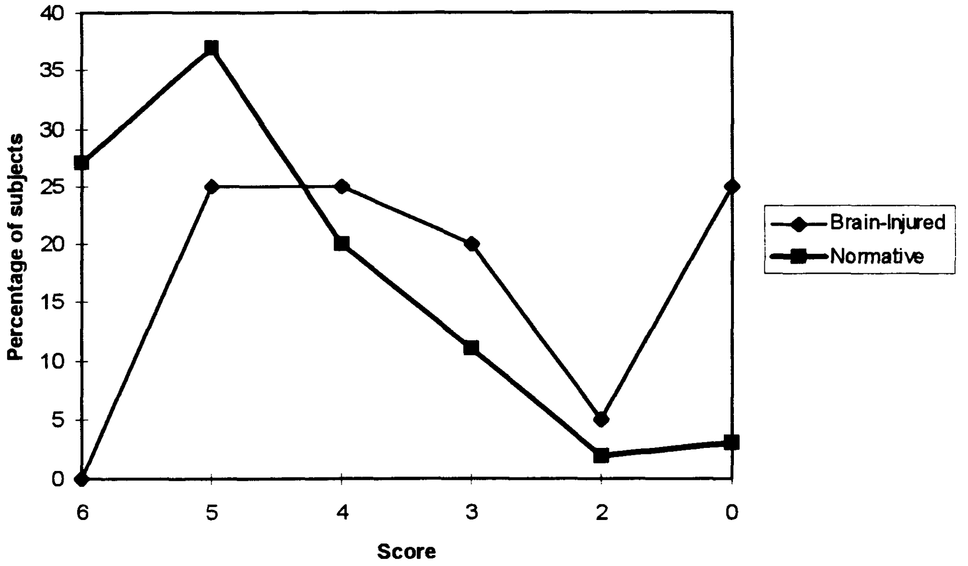

*Figure 44: Comparison of brain-injured and normative groups on Seals 3-ball problem scores as a function of trials to criterion.*

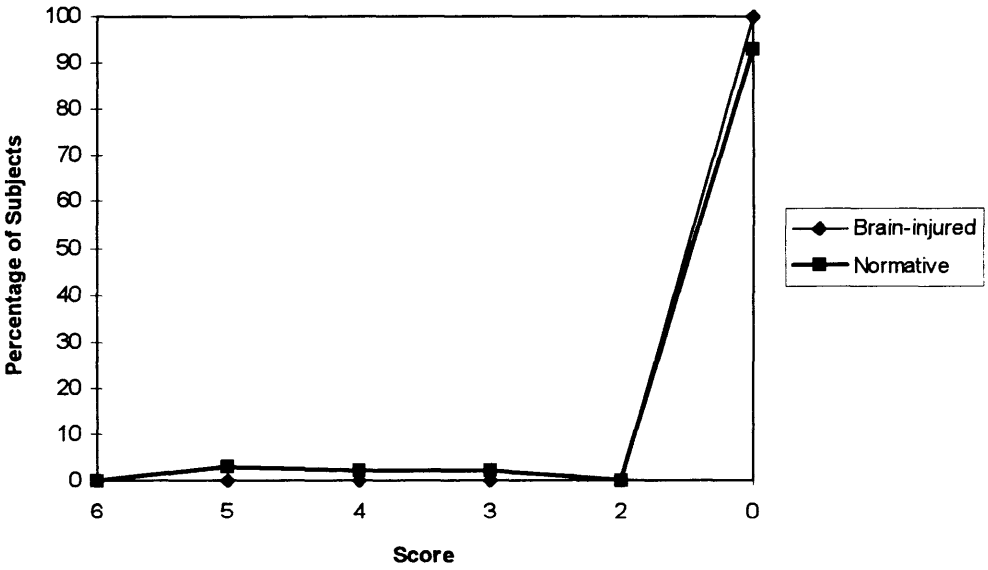

*Figure 45: Comparison of brain-injured and normative groups on Seals 4-ball problem*

It is evident, from Table 74, that 27% of the normative sample successfully solved the 3-ball problem in the minimum number of trials, scoring 6 points, whereas not one of the brain-injured subjects was able to do so. In addition, 97% of the normative sample was able to successfully solve the 3-ball problem within the maximum six trials, while only 75% of the brain-injured subjects was able to do so. Both groups found the 4-ball problem a difficult task with very few subjects able to solve the task within the six trials. For example, none of the brain-injured subjects was successful on this task, whereas 7% of the normative subjects were able to solve the 4-ball problem within the six trials.

### Error-Types and problem-solving Strategies

The key errors for the 3 and 4-ball problem tasks were identified by examining the most frequently occurring errors for the group. As for the normative sample, the pattern of errors for the brain-injured group 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. Only Trials 1 and 2 were examined as this ensured that all subjects' results would be counted. Recall that if subjects were successful on Trials 1 and 2, then they were not presented with further trials for that problem type.

For both the 3 and 4-ball problems, the majority of errors occurred on Move #1. For the 3-ball problem, 95% of subjects made the first error on Move #1, and for the 4-ball problem 50% of subjects made their first error on Trial #1. On Trial 2 of both the 3 and 4-ball problems, the majority of errors occurred on Moves #3 and #5. This is a similar error pattern to the normative sample, where the majority of errors occurred on Moves #1, #3, #5, and #9.

An analysis of subjects' verbalisations during performance on the Seals task also reveals a striking similarity with those results from the normative sample. On both 3 and 4-ball problems, the vast majority of verbalisations occurred during those same moves identified as key error points in the task for the normative sample. Anecdotal information was recorded for thirteen of the twenty subjects. The remaining seven subjects did not offer any verbalisations during testing. A verbalisation was defined as any utterance that occurred during performance on a particular move. It ranged from a single word to a string of two or three sentences. This was categorised as one instance of verbalisation. All verbal responses were recorded verbatim. On the 3-ball problem, nine instances of verbalisation were recorded over the twenty subjects (this represented only the first verbalisation for each subject). Of those nine verbal responses, three related to Move #1, four related to Move #3, one related to Move #4, and one related to Move #5. Eight of the nine verbalisations occurred during a response to the key error points of Move #1, #3 and #5 which had been identified earlier. Subject #518, for example, during Move #1 of the 3-ball problem made the following comments: "Can you move three at a time?" A few seconds later, while still looking at the screen: "Can you move from the bottom of the stack?" Subject #512 during Move #3 of the 3-ball problem asked: "Can you move it back?" (referring to moving the green ball back on top of the red ball).

During performance on the 4-ball Seals problem, ten instances of verbalisation were recorded. Five instances occurred on Move #3, one during Move #4 and four occurred during Move #5. Nine of the ten examples occurred during the key error points. Some examples of these are: Subject #515 during Move #3 "You can't put the blue on top of the red." Subject #500 during Move #3 "Can't put blue on top of red, that's against the rules". Subject #505 during Move #1: "Can you move a ball from the bottom of the pile?"

It is apparent, as with the normative sample, that the vast majority of subjects' verbalisations occurred during those critical choice or error points in the task.

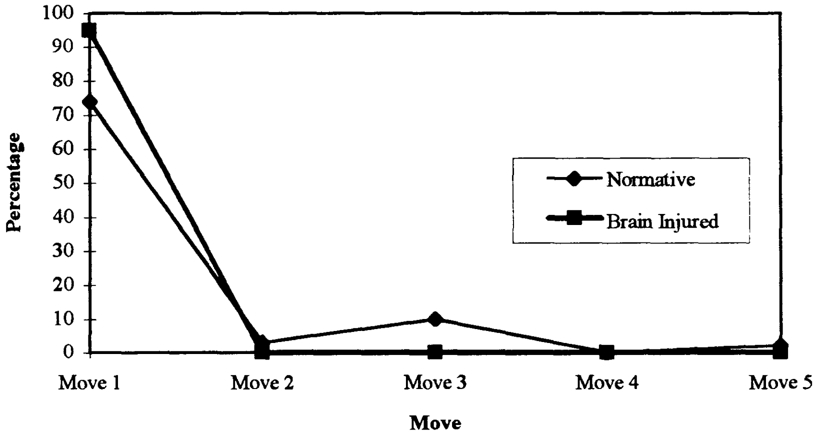

***Figure 46: Comparison of percentage of errors of normative and brain-injured subjects on Trial 1 of 3-ball Seals Task.***

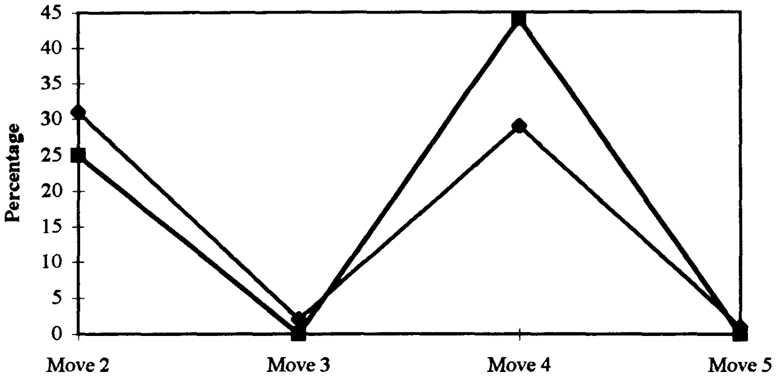

*Figure 47: Percentage of errors for normative and brain-injured groups as a function of moves on Trial 2 of Seals 3-ball problem*

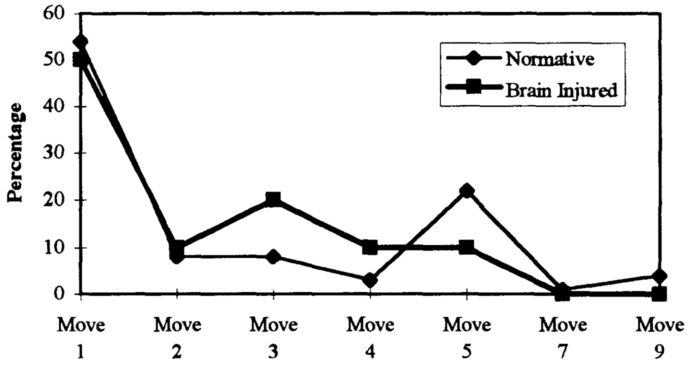

*Figure 48: Percentage of errors for normative and brain-injured groups as a function of moves on Trial 1 of the Seals 4-ball problem*

### 8.3.3 Balance Beam Task

Mean score for the Balance Beam for the brain-injured sample was 63.8%, compared with 67.5% for the matched normative sample. While this did not represent a statistically significant difference (F=1.369, p=0.246), the trend is for lower scores for the clinical group. There is little difference in standard deviations between both groups (10.6, 10.5), suggesting comparable variability in performance. A Cronbach alpha internal reliability coefficient of r=0.626 was obtained using Testat. This compared with r=0.561 for the normative sample. The Balance Beam is considered to have an acceptable level of internal reliability.

### 8.3.4 Fish Task

Mean scores for Fish were 90.4%, compared with 91.8% for the normative sample. There is clearly little difference in scores between the two groups (F=0.210, p=0.648). There is a larger range of scores for the brain injury group, with a standard deviation of 13.4, compared with 11.4 for the normative sample.

### 8.3.5 Austin Maze Scores

**Table 75: T-tests for significance between normative and brain-injured groups on Austin Maze measures.**

| | Total Errors | Mean Errors | Time Taken | Error 1 | Error 2 | Error 3 | Error 4 | Error 5 |
|---|---|---|---|---|---|---|---|---|
| T-score | -0.913 | 1.917 | 1.392 | -1.538 | 0.050 | -1.167 | 2.173 | -0.744 |
| p-value | 0.373 | 0.171 | 0.242 | 0.141 | 0.960 | 0.259 | 0.043* | 0.466 |

* significant at $p < 0.05$ confidence level.

The results of Table 75 indicate that only one Austin Maze measure, Error Type 4, separated the two groups at a significant level. Recall that Error Type 4 is incidence of illegal diagonal moves. A somewhat surprising result is that the normative subjects incurred more of these illegal moves than the brain-injured group, with mean scores of 2.709 and 1.263 for the normative and brain-injured subjects respectively. Failure to inhibit an illegal move is more likely to be a consequence of the brain-injured than the normative subjects. One possible explanation may be that the normative subjects focused their attention more on the computer screen than on the keyboard in their attempts to recall the correct pathway, and in so doing, failed to take account of the positions of the four correct keys. It was observed, for example, that a number of the normative subjects were surprised to find that they had pressed the wrong key. This observation was not formally tested.

A *post hoc* comparison of the Austin Maze Time Taken scores indicates that, as a group, the brain-injured subjects took longer to complete the Austin Maze Test (mean of 936.6 seconds, compared with 851.4 seconds for normative group). However, this difference is not significant (t= 1.392, p=0.242). For both groups, there is considerable variability in scores, with a standard deviation of 335.6 for the brain-injured and 311.6 for the normative sample.

**Error-Utilisation**

A comparison of subjects' scores on the first ten trials was made with total errors to determine to what extent these were predictive of scores on later trials. A Pearson correlation coefficient was derived from the two scores, and the results are r = 0.948. This result compares with a correlation of r = 0.955 for the matched normative sample, suggesting that for both groups, scores on the first ten trials are a good estimate of overall errors.

### 8.3.6 WCST

The results of the comparison of performance of the normative and brain-injured subjects on the WCST indicate that the two error measurements clearly separated the performance of the two groups. The results are detailed in Table 76. Both Perseverative Errors and Total Errors on this test were discriminators that separated the performance of the brain-injured and the normative subjects. The normative subjects made significantly less errors and less perseverative errors than the brain-injured subjects. There was also a difference in the number of Correct Sets obtained, with the normative subjects producing a higher score, although this difference was not significant at the $p < 0.05$ confidence level.

**Table 76: Comparison of performance of normative and brain-injured subjects on WCST measures**

| | Time Taken | Correct Sets | Perseverative Errors | Total Errors |
|---|---|---|---|---|
| Normative (mean scores) | 349.875 | 3.429 | 6.946 | 17.411 |
| Brain-injured (mean scores) | 354.263 | 2.632 | 9.579 | 24.263 |
| t-score | 0.017 | 1.966 | -3.305 | -3.341 |
| p-value | 0.987 | 0.065 | 0.004* | 0.004* |

* significant at $p < 0.01$ level

There is no significant difference between the two groups in the amount of time taken to complete the WCST (t=0.017, p= 0.987). A direct comparison of time scores indicates less variability in performance of the brain-injured sample (SD = 71.6) compared with the normative group (SD = 83.5).

### 8.3.7 AVLT

There is no significant difference between the two groups on the seven measures of the AVLT. A comparison of independent t-tests for the brain-injured and the normative sample is detailed in Table 77. The results suggest that the brain-injured and normative groups did not differ in their performance on a measure of short-term memory and learning of a string of words. A comparison of the learning curves of the acquired brain-injured and normative groups over the five learning trials is detailed in Figure 49.

**Table 77: T-tests of scores on AVLT measures for brain-injured and matched-age normative groups.**

| | Retroactive Interference | Proactive Interference | Forgetting | Learning | Retrieval | Recognition A | Recognition B |
|---|---|---|---|---|---|---|---|
| T-score | 0.792 | -0.719 | -0.764 | 1.666 | 0.617 | 0.156 | 0.156 |
| p-value | 0.438 | -0.481 | -0.454 | 0.201 | 0.545 | 0.878 | 0.878 |

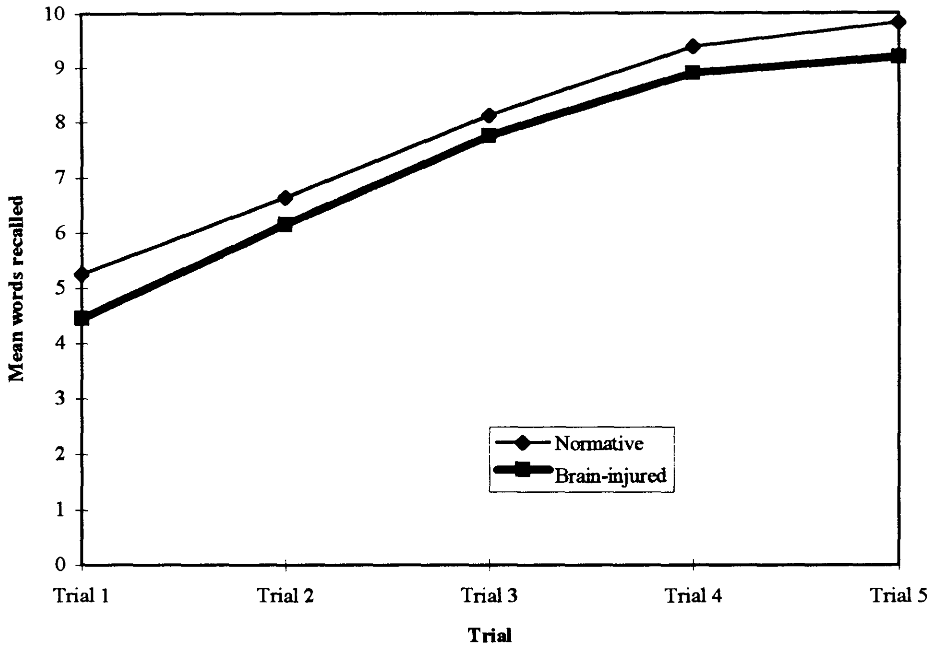

*Figure 49: Comparison of normative and brain-injured groups on number of words recalled for each of the five learning trials on the AVLT*

### 8.3.8 OLSAT Total Score

The mean OLSAT Total score for the brain-injured sample is 86.3 (SD=12.9). There is no significant difference between this result and that for the matched normative group of children (88.9, SD = 16.3; F=0.053, p=0.818). Both groups of subjects scored in the low average level of school ability.

## 8.4 Comparison of Brain-injured sample with Junior Normative Sample

It has been demonstrated that there was a significant difference in scores between the brain-injured group and matched normative group on the Seals task. The question now arises whether the brain-injured group performed more like the Junior 7-8 year old non-clinical age group than with their matched aged peers. This comparison was examined in the context of the wider issue of comparison of scores across all developmental tests. T-tests were conducted on each of the pairs of scores and Table 78 details the outcomes.

**Table 78: Comparison of mean scores(s.d. in brackets) and tests of significance for Brain-injured and Junior Group on developmental measures**

| Tests | Brain-injured sample | Junior Group | T-test | p-value |
|---|---|---|---|---|
| Seals | 46.9 (13.9) | 48.3 (13.1) | -0.664 | 0.515 |
| Balance | 63.8 (10.6) | 62.9 (9.65) | -0.717 | 0.482 |
| Fish | 90.4 (13.4) | 88.6 (15.25) | -0.525 | 0.603 |
| Piggy Bank | 89.5 (11.0) | 87.6 (13.9) | -1.964 | 0.064 |

The results of Table 78 suggest that for the four developmental measures there is no significant difference in scores between the brain-injured and the Junior age group of children. Taken together with the outcome detailed in Table 73, this would suggest that the brain-injured group performed more like the Junior Group than the matched normative age group. There are several points of evidence for this conclusion. A *post hoc* inspection of the mean scores for the four measures indicates that the brain-injured group's mean scores is closer to the Junior than the Middle normative group. For the Seals Task, in particular, the evidence is statistically stronger in that there is a significant difference between the brain-injured and normative sample, but not for the brain-injured and Junior sample (F=16.817, p=0.000 for Normative, and T= -0.664, p=0.515 for Junior Group). An inspection of Figure 50 indicates that the graphs depicting scores for the Junior and the brain-injured group are closely aligned.

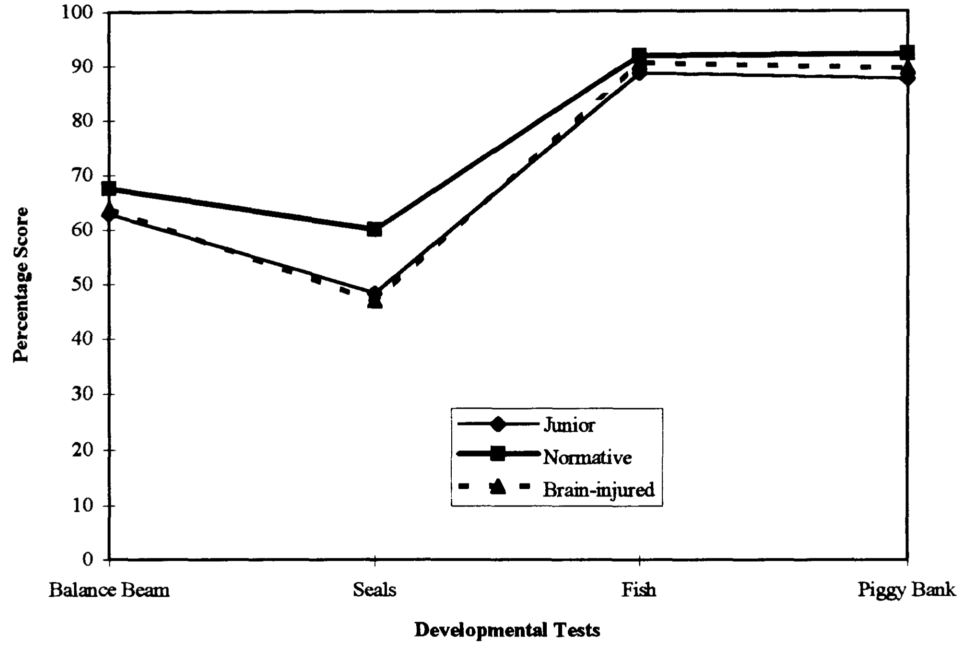

***Figure 50: Comparison of Brain-injured, Normative 9-10 year group and Junior 7-8 year age groups on four experimental developmental tasks***

## 8.5 Canonical Analysis

A Canonical Analysis between the brain-injured and matched-age normative groups was computed using Systat. The dependent variables of Seals and Balance Beam, the two developmental measures of executive functions, were hypothesised to dominate the loadings on the canonical function that separated the clinical and normative samples. It has already been confirmed that Seals and Balance Beam positively dominated the loadings on the Executive Function factor when employed with a normative sample of children. This result is detailed in Table 60.

Ten variables were used in the canonical analysis. The variables were: Austin Maze mean errors, Austin Maze Time taken, AVLT number of words remembered, Seals, Balance Beam, OLSAT Total Score, WCST number of correct sets, WCST time taken, Fish and Piggy Banks. It is clear from Table 76, which details the canonical correlations, that the Seals task is the only measure which significantly provides a separation between the clinical and normative group for this age range of subjects. The canonical correlation for the Seals task is 0.853 which is a significant result. The number of correct sets for the Wisconsin Card Sorting Test approaches significance with a correlation of 0.407. A *post hoc* comparison of mean scores between groups indicates comparable results on the OLSAT and the number of correct words remembered on the Auditory Verbal Learning Test.

The Canonical correlation is 0.518, which is a significant result. The separate dependent variable canonical loadings confirms that, of these dependent variables, Seals accounts for the majority of the variance on the group variable with a correlation of 0.853 with the canonical value.

The two Austin Maze measures, the AVLT score and OLSAT score have moderately low canonical loadings suggesting they add little to the variance in explaining the canonical function. The separate loadings are detailed in Table 79.

Given the magnitude of the Seals test statistic, then how significant a discriminator is it on its own? The Canonical correlation of 0.518 sets the upper limit for any correlation that separates the two groups. A Pearson correlation matrix involving the ten variables reported in the canonical analysis as well as a group variable (1=normative sample, n=60 and 2=brain-injured sample, n=20) was conducted between the normative and clinical samples using Systat. The correlation between Seals and the Group variable is r= 0.459 (see Table 79 for the Group variable results), which approaches the canonical correlation, suggesting that on its own, the Seals task is a good predictor of brain injury involving children in the 9-10 year age range. The nearest correlation is WCST Correct, with a correlation of r=0.239.

**Table 79: Canonical correlations and coefficients, and Pearson correlations with group variable for acquired brain injury and normative groups.**

| | **Canonical Correlations** | **Coefficients** | **Group** |
|---|---|---|---|
| Ausmeane | -0.288 | -0.269 | -0.173 |
| Austime | -0.106 | 0.227 | -0.064 |
| AVLT | 0.269 | 0.191 | 0.161 |
| Balance | 0.243 | -0.006 | 0.146 |
| OLSAT | 0.048 | -0.125 | 0.029 |
| Seals | 0.853 | 0.882 | 0.459 |
| WCST correct | 0.407 | 0.331 | 0.239 |
| WCST time | 0.038 | 0.405 | -0.023 |
| Fish | 0.101 | | 0.123 |
| Piggy Bank | 0.221 | | 0.201 |

## 8.6 Sensitivity, Specificity and Predictive Power of Tests

The sensitivity and specificity of the Seals, WCST and the Austin Maze were separately analysed and compared, using the formula set out in Elwood (1993, p227), and reproduced below in Table 80. The sensitivity of a particular measure refers to the proportion of subjects with a brain injury who are correctly identified by the selected measure, whereas specificity refers to the proportion of subjects who do not have a brain injury who are successfully excluded by the measure. According to Elwood (1993), neuropsychologists are beginning to see the value in specifying these statistics.

**Table 80: Classifications and Probabilities for Specificity and Sensitivity**

|  | + | - |  |
|---|---|---|---|
| + | a | b | (a + b) |
| - | c | d | (c + d) |
|  | **a + c** | **b + d** | **N** |

a= true positives
b= false positives
c= false negatives
d= true negatives

$\frac{a}{a+c}$ = sensitivity

$\frac{d}{b+d}$ = specificity

$\frac{a}{a+b}$ = PPP (Positive Predictive Power)

$\frac{d}{c+d}$ = NPP (Negative Predictive Power)

### 8.6.1 Seals Task

Applying the formula in Table 80 to the Seals task, comparing the sixty normative 9-10 year olds with the 20 brain-injured 9-10 year olds, produces the following figures, detailed in Table 81:

**Table 81: Bayesian derivations for Sensitivity, Specificity and Predictive Values for Seals Task**

|  | + | - |  |
|---|---|---|---|
| + | **9** | **11** | **20** |
| - | **4** | **56** | **60** |
| **N** | **13** | **67** | **80** |

Table 81 indicates that when the cut-off point for the Seals task is set at 49% accuracy, nine of the twenty brain-injured subjects are correctly diagnosed as being positive for brain injury (eleven subjects are false negatives). In addition, four of the normative sample of sixty subjects are false positives (falsely identified as having brain injury), whereas fifty-six of the normal subjects are correctly excluded by the test as not having brain injury. Using the equations set out in Table 80 produces the following results for the Seals test:

| | |
|---|---|
| Sensitivity: | 0.7 |
| Specificity: | 0.84 |
| Negative Predictive Power: | 93% |
| Positive Predictive Power: | 45% |

It is apparent from the above statistics, that the Seals has respectable Negative Predictive Power and specificity. When applied to a population of normal children enrolled in regular classes, the test correctly predicts, with 93% accuracy, that a negative score reflects intact brain function.

### 8.6.2 WCST

The WCST Correct Sets variable was used for the comparison, as it loads significantly onto the Executive Function factor. Using a cut-off point of $\leq 2$, Table 82 shows the number of subjects from the brain-injured and normative 9-10 year groups who fall within the cut-off point.

**Table 82: Bayesian Derivations for Sensitivity, Specificity and Predictive Values for WCST**

| | + | - | N |
|---|---|---|---|
| + | **9** | **10** | **19** |
| - | **13** | **43** | **56** |
| **N** | **22** | **53** | **75** |

It can be seen from Table 82 that when the cut-off point of $\leq 2$ is set, nine of the nineteen brain-injured subjects are correctly identified as having a brain-injury, and forty-three of the sixty normative 9-10 year old subjects are correctly excluded from having a brain-injury.

Using the formula from Table 80, produces the following statistics for the WCST:

| | |
|---|---|
| Sensitivity: | 0.41 |
| Specificity: | 0.81 |
| Negative Predictive Power: | 77% |
| Positive Predictive Power: | 47% |

The WCST has weak sensitivity. With a Positive Predictive Power at less than chance level, and a Negative Predictive Power that would leave the clinician with a 23% chance of misdiagnosis, the WCST would not be a good measure of choice for predicting brain damage in a group of 9-10 year old children.

### 8.63 Austin Maze

The Austin Maze Mean Errors per trial variable was used in this analysis as it loads significantly onto the Executive Function factor. A cut-off point was set at $\geq$6.5 errors per trial to maximise predictive power. From Table 83 below, and using the formula established by Elwood (1993), the following measures were derived.

**Table 83: Bayesian Derivatives for Sensitivity, Specificity and Predictive values for Austin Maze**

| | + | - | N |
|---|---|---|---|
| + | **9** | **10** | **19** |
| - | **20** | **38** | **58** |
| **N** | **29** | **48** | 77 |

| | |
|---|---|
| Sensitivity: | 0.31 |
| Specificity: | 0.79 |
| Negative Predictive Power: | 67% |
| Positive Predictive Power: | 47% |

It is apparent that the Austin Maze has weak sensitivity and specificity. It correctly identifies nine of the brain-injured subjects, and correctly excludes only thirty-eight of the fifty-eight 9-10 year old normative subjects, misdiagnosing 33%. The Austin Maze has Positive Predictive Power at less than chance level, indicating that it correctly identifies less than half of the brain-injured subjects. In addition, its Negative Predictive Power, in correctly excluding the normal subjects, does not instil confidence in a clinician, with only 67% accuracy. This result means that, with a group of normal children, a negative score on the Austin Maze can be taken to preclude the possibility of brain damage with only 67% certainty. There is a 33% possibility of a false negative.

In summary, the results of Tables 81-83 indicate that on an individual case basis, the Seals test has the greatest level of sensitivity and specificity in comparison to the Wisconsin Card Sorting Test and the Austin Maze. This result is summarised in Table 84 below. On the basis of the current results for 9-10 year old children, both the WCST and the Austin Maze do not inspire confidence that either test is sensitive or specific to brain injury. It is suggested, however, that the Seals task has a respectable level of specificity (0.84), although would be considered to have only modest sensitivity (0.7). In addition, the Seals task has a significant level of Negative Predictive Power with 93% accuracy. In diagnostic terms, what does this result mean?

**Table 84: Sensitivity, Specificity and Predictive Power of Tests**

| | SEALS | WCST | AUSTIN MAZE |
|---|---|---|---|
| SENSITIVITY | 0.7 | 0.41 | 0.31 |
| SPECIFICITY | 0.84 | 0.81 | 0.79 |
| NPP | 93% | 77% | 67% |
| PPP | 45% | 47% | 47% |

According to desRosiers (1992a, p310), tests which have good specificity and Negative Predictive Power, have clinical utility as "second-stage tools". In a clinical diagnosis, highly specific tests are vital when they produce abnormal results in that they give a strong indication that the condition is actually present and not due to some other pathology. A clinician would look for tests which are highly sensitive to brain-injury as the first-line measure, and would then, as a second stage of diagnosis, use tests which are highly specific to brain-injury, to help confirm the presence of brain damage.

Consider the case of a child referred to a neuropsychologist because he had suffered a blow to the head as a result of a fall, and his schoolwork had begun to deteriorate. The clinician would query brain damage. Suppose the neuropsychologist chooses an instrument which has high sensitivity and Positive Predictive Power of 94% for discriminating brain injury for the particular population of children in question. This is usually a first screening test. If the test is positive, there is still a 6% chance of misdiagnosis. The neuropsychologist concludes from the positive result of first screening that there is some abnormality in the child's performance. However, the child may have a specific learning difficulty, or may be suffering from anxiety which has depressed his performance. At this point, there is a need to further narrow down the cause of the decrement in performance. A test which has high specificity and Negative Predictive Power for brain injury is then selected to confirm that the child's decrement in performance is only due to brain damage. This second-stage testing would be critical in the diagnostic process to avoid the possibility of type-1 errors. It is contended that the Seals test is a measure which has good specificity for indication of brain-injury in a normal population of children, and, in the current context, would be recommended over the WCST and the Austin Maze for second-stage diagnosis of brain damage.

There is a further issue which needs to be raised at this point. It has been argued, and evidence presented, attesting to the high Negative Predictive Power of the Seals test, in comparison with the adult neuropsychological tests. Evidence has also been presented attesting to the low Positive Predictive Power of the WCST and the Austin Maze, and the moderate Positive Predictive Power of the Seals test. The sensitivity of the tests is based on the premise that they are predictors of brain damage in a group of heterogeneous subjects. However, the line taken in this study has been that these tests are measures of prefrontal lobe functioning, and not generic tests of brain damage. For example, of the twenty children with brain damage used in this study, 19% had frontal damage confirmed by CT scans. Given that the Seals test was not designed as a generic measure of brain damage in children, then the result that it obtained a Sensitivity score of 0.7, heightens its clinical utility in identifying brain damage in a population of children.

# 9.0 DISCUSSION

## 9.1 Underlying Factor Structure

This study examined the performance of three age groups of young males on four experimental tests of executive function which were developed according to Piagetian and neuropsychological precepts. The three age groups were 7-8 years (n=62), 9-10 years (n=60) and 11-12 years (n=61). In addition, the same groups were tested on adult measures of neuropsychology, namely the WCST, the Austin Maze and the AVLT, as well as the OLSAT which is a measure of school ability. The main finding is a clear developmental sequence for some tests but not for others. There appears to be a clear parallel between improvement in performance on neuropsychological tests and Piagetian stages of cognitive development. In the following discussion the data of the current study are reviewed in light of this conclusion.

Analysis of the results showed that for a non-clinical sample of 183 healthy boys between the ages of 7 and 12 years, who were administered three neuropsychological measures of executive function, experimental developmental tasks based on Formal Operational Reasoning and a school ability test, there are three underlying factors which account for the data. These factors were labelled Executive Function, Memory/Learning, and Inhibition. The Executive Function factor is defined by the Seals Task, Balance Beam, WCST Correct score and two Austin Maze variables (Time taken and Mean Errors per trial). The Memory/Learning factor is defined by the OLSAT Verbal and Nonverbal variables, an AVLT Learning variable and WCST Time. The Inhibition factor is defined by Fish and Piggy Bank, both of which involve inhibition of a previously learned or inappropriate response.

Pair-wise comparisons of the correlations of each underlying factor for each age group, using Bonferroni adjustments, indicate that there is a significant difference across each of the three age groups for the first Executive Function factor. This robust result confirms that an Executive Function Factor follows a developmental sequence across the age ranges employed in this study. The second Memory/Learning and third Inhibition factors are not as impressive in their separation of the three age groups along developmental lines. A MANOVA reveals that the Executive Function Factor accounts for the majority of variance (0.922) of the Canonical function that defines the entire sample of 183 children by age.

In addition, for the Executive Function and Inhibition factors, the standard deviation scores decrease with age suggesting less variability in scores with increasing age. This result indicates a greater stability of performance with increasing age, and reflects the notion that with the first and third factors, there is an increasing homogeneity in function with increasing age.

Pearson correlations among the three factor scores indicate an increasing correlation of factors with increasing age. Correlation coefficients range from r=0.309 to r=0.596. The results of the Pearson correlation coefficients suggest two possible implications. First, the significant correlations between the first factor, Executive Function and the second factor, Memory/Learning, are viewed as consistent with the hypothesis that as executive function develops, so does memory. The proposed CN Model asserts that executive function is defined as the interaction between active memory and planning processes and a significant correlation, as demonstrated in the findings, between memory and executive function is consistent with the model. The result provides support for the model. These findings tend to confirm the notion that "memory is embedded in executive function" (Welsh, Pennington and Groisser, 1991, p 144 ).

The second implication stems from the outcome that there is an increasing correlation of factors with age, particularly between the first Executive Function and second Memory/Learning factors. This result suggests that there is a greater independence of factors for the youngest age group. The results indicate that with increasing age, there is a growing relationship between executive function and memory performance and it is contended that this outcome is consistent with the CN Model proposed in this study. At a cognitive level, the result suggests that memory becomes increasingly more important in the performance of problem-solving activities that involve planning processes. This phenomenon is a function of the complexity of tasks leading to the attainment of Formal Operational Reasoning. The younger Concrete Operational Reasoner appears to examine each piece of information, singly and in isolation. Decisions are made on the basis of salient features rather than taking all the relevant information into consideration (Ginsburg and Opper, 1988). However, the Formal Operational Reasoner, around 11-12 years of age, appears to take account of all information, weighing each in relation to the other, and deducing an outcome. If that outcome does not eventuate, then another strategy is developed, based on other possible permutations. The implication that a greater reliance on active memory processes with increasing age is consistent with the present outcome, is explainable in Piagetian terms. VanLehn (1991) analysed the moves of an 11 year old subject performing the Tower of Hanoi where no feedback on performance was given. He concluded that the subject's "overall approach to strategy acquisition method appears to be the classic 'scientific method' of hypothesis formation and experimentation" (VanLehn, 1991, p16). VanLehn was describing hypothetico-deductive reasoning which involves not only the development of hypotheses, testing of these in practice and evaluation of their performance, but the ability to then return to the original plans and generate a new hypothesis. According to Piaget (1976), the ability to return to the original plan and then generate a new strategy involves the process of 'reversible mental operations' which occurs during the attainment of Formal Operational Reasoning. VanLehn's subject, for example, deliberately chose a different first move on the second trial after she became stuck on the first trial of the Tower of Hanoi. She apparently remembered the sequence of moves during the first trial and was able to identify her first move. The involvement of active memory in the process of reversible mental operations becomes increasingly more important in maintaining the sequence of events or moves so that a different plan can be instituted. The Seals Task of the present study, in the active memory mode, places considerable memory demands on a subject's ability not only to use planning and hypothetico-deductive reasoning, but to hold the proposed moves in active memory while planning occurs. In the case of the youngest boys in the sample studied here, it is acknowledged that their memory capabilities are not as advanced as those of the oldest children, and, as a result, have less impact on the performance of problem-solving activities. A caveat should be placed here. The present study was not designed to tease out the separate influences of executive functioning and memory. However, the results of correlational coefficients between the first and second factors would suggest that a positive and developing relationship between the two does exist with increasing age.

An analysis of the individual tests which load onto the Executive Function factor indicates that only the Seals test provides a clear developmental progression, as reflected in improved scores, across the three age groups. Subjects displayed a significant improvement in scores, as the age range moved from the Junior through to the Middle to the Senior age groups. Scores on the Balance Beam, on the other hand, displayed significant improvement between the Junior and Senior groups and the Junior and Middle groups, but not Middle and Senior groups. Balance Beam scores plateaued between the Middle and Senior age groups, with a ***post hoc*** analysis indicating little difference in scores between the 9-10 year group and the 11-12 year group, suggesting that for this particular sample, performance on the Balance Beam reaches asymptote by the age of 9-10 years. The WCST correct number of sets scores, another variable which loaded onto the Executive Function factor, displays a two-stage level of maturation across the three age groups. The test is not sensitive to developmental changes below the age of 9-10 years. This point is taken up in Section 9.12.

Of the two Austin Maze variables which load onto the Executive Function factor, Mean Errors per Trial and Time Taken, Mean Errors display a significant improvement, between the Junior and

Senior groups and the Middle and Senior groups, but not the Junior and Middle groups This result suggests that significant improvement in performance occurs around 9-10 years of age. From the age of 11-12 years, scores then begin to resemble those of adults (Hume, 1986). As with the WCST, the Austin Maze displays a two-stage level of maturation, failing to discriminate the performance of children below the age of 9-10 years. This point will be elaborated in Section 9.10.

The underlying Memory/Learning factor was defined by the variables of WCST Time, OLSAT Verbal and Nonverbal, and AVLT Learning. It is contended that all four variables have a significant memory component, with the AVLT being a test of memory *per se*. Three of the four variables are tests which are timed (the AVLT being the only exception), suggesting that speed of responding may also impact on this factor. In fact, the AVLT did have its own internal timing in that words are read to all age groups at the same rate of one word per second. This imposition of an internal timing factor from the AVLT, for all age groups, places a structure upon the child's cognitive processing and suggests that speed of processing may also be a factor affecting performance on the AVLT as there is a differential performance effect with age. It is noted that both experimental developmental executive function tests, the Balance Beam and the Seals task also load onto this factor, although their loadings are considered not as significant as the other variable loadings on the Memory/Learning factor (0.410 and 0.423 respectively). However, both Seals and Balance Beam do have an active memory component, and it is not surprising that they should load onto a separate Memory/Learning factor in addition to the Executive Function Factor. This result, of course, strengthens the argument covering the relationship between executive function and memory.

The third underlying factor, that of Inhibition, is defined by Fish and Piggy Bank, both of which have highly significant factor loadings (0.717 and 0.720 respectively). Both tests have a major component of response inhibition. Subjects must make a decision, not necessarily based on the previous response. That is, they may have to inhibit a previously-learned response in order to be correct. Both tests involve visual distractors which subjects must inhibit in order to respond appropriately. This underlying third Inhibition factor is clearly defined by these two tests, while all other tests have poor or nil loadings.

The results of the three factor solution of the present study provide support for the notion that measures of Executive Function, based on Piagetian Formal Operational Reasoning, yield results which are sensitive to developmental improvement in performance. This is a central part of this thesis and provides support for the first hypothesis of the study, that executive functions follow a developmental sequence of improvement. There is a significant improvement, across the three age groups, on the Executive Function factor confirming its developmental timetable. The Executive Function factor is the only one of the three identified factors which clearly separates the three age groups. The two developmental measures, Seals and Balance Beam, load significantly onto the Executive Function factor. Seals on its own, also provides a significant developmental improvement in subjects' scores across the three age groups, confirming its status as a measure which clearly has good discriminant properties for children in the 7-12 year age category. It is argued that the major contributing factor from this test is the inclusion of the 'active memory' component and the developmental nature of the measure. That is, the Seals test is graded in difficulty, beginning with tasks that are at a simple level and progressing to tasks that are more complex. For example, the Seals task begins with the 2-ball problem, where solution is achieved in a minimum of three moves and progresses to the 4-ball problem where a solution is achieved in a minimum of fifteen moves.

The Seals task in the active memory mode, used in the current study, unlike the traditional presentation of TOH-type tasks reported in other studies, may well be a critical factor in producing the significant discrimination in performance across the three age groups. However, this assertion was not directly tested. When the TOH was given in the traditional presentation to children in the age ranges of 3 to 12 years, where subjects were able to manipulate the pieces and given immediate feedback on their performance, there was improved performance for the youngest age group compared with results reported previously in the literature (Welsh, 1991). The differences between the age groups were minimised.

The present results provide support for Piagetian stages of cognitive development, with 7-8, 9-10 and 11-12 years of age being periods where significant growth in mental development occurs. The parallel between the growth and development of cognitive and executive functioning in children within these age groups is supported. The Executive Function Factor, provides a clear separation between the three age groups and further support for the notion that paediatric measures of executive function, based on Piagetian cognitive development, are valid test discriminators.

## 9.2 CN Model

On an individual test basis, do the current results provide support for the CN Model in the way that it is used in this study? The CN Model was developed as an extension of Piagetian theory of cognitive development to take account of the coincidental data on neural development. The CN Model is perceived as an explanatory model and the results of the present study are in no way to be taken as a test of the model. A test of the model would necessarily involve neurophysiological and neuroanatomical data, and would also address the issue of whether a dissociation can be established between assimilation and accommodation, in the same manner that a dissociation can be demonstrated between active memory and planning processes. The CN Model was used as a theoretical basis for the development of paediatric measures of executive function. Table 2 in Section 2.11 outlines the predictions that were made for each of the developmental and neuropsychological tests. These predictions specified that the Seals and the Balance Beam tests would be the only two predicted to be paediatric measures of executive function. Evidence has been presented which clearly supports the Seals test as the only measure used that significantly separated the three normative age groups **as well as** the 9-10 year old brain-injured subjects from the matched normative subjects. In Section 8.3.2 it is contended that the results of the Seals test provide support for the CN Model in that successful solution to the task involves an interaction between active memory and planning skills and the measure is based on developmental skills. Although the Seals test had modest sensitivity (0.7) and acceptable specificity (0.84), its Negative Predictive Power in excluding a normal group of children from brain injury was 93%. These statistics support the clinical utility of using the Seals test as a measure of executive function with children.

The predictions from the CN Model, however, specify that the Balance Beam test would be an effective paediatric measure of executive function. It was predicted that the Balance Beam would be sensitive to developmental changes and predictive of brain-injury in a paediatric population. The results, however, do not support this prediction. The task analysis, detailed in Table 2, indicates that the Balance Beam contains components of active memory and planning, pre-requisites for executive functioning, as espoused by Fuster (1993). In addition, the Balance Beam test is based on developmental psychology, contains a graded level of difficulty, and was predicted to be sensitive to developmental changes in cognitive development. This is not the case.

It is contended that in the review of the task analysis, detailed in Table 2, the Balance Beam task in its present form, contains a number of complexities which would preclude it from being a strong measure of paediatric executive function. Certainly, the planning skills aspects of the task can be well demonstrated. The task involves hypothetico-deductive reasoning that takes account of more than one dimension, inhibition of a response that would be based on just the number of disks on either side of the fulcrum, and, the use of environmental feedback, based on the way the beam moves, so that responses can be appropriately adjusted.

However, it is the active memory component of the Balance Beam test, which may have to be brought into question. The task analysis, involving an active memory component, was made on the basis that subjects hold in active memory, previous responses as well as a repertoire of strategies, for solving the task. For example, possible strategies are: count the number of disks on either side of the fulcrum, count the number of pegs from the fulcrum, look at both distance from the fulcrum and number of disks, and multiply the number of disks by the number of pegs from the fulcrum. Does this concept of active memory when solving the Balance Beam equate with the active memory component in solving the Seals test? There are differences. On the Balance Beam test, subjects are given feedback on their performance. They decide on one of three possible alternatives and then watch the beam fall to see whether they are correct. This is not the case with the Seals test, where subjects hold moves in memory and do not receive feedback on their performance. Allowing subjects to receive feedback and monitor their performance on the TOH changes the task demands such that younger children are able to successfully solve the task (Welsh, 1991). This point is considered further in Section 9.18.

In addition, subjects are told the rules of the Seals task. The three rules for solving the task are explicitly stated, and subjects are given practice in using the rules on the 2-ball problem. A rule violation is brought to the subjects' notice on the 2-ball problem. On the other hand, subjects have to **discover** the rules for solving the Balance Beam task. Although feedback on performance is given in the Balance Beam, subjects must use the information to discover the rules or strategies for solving the task. This aspect adds considerably to the complexity of the Balance Beam task.

A *post-hoc* analysis of the Balance Beam task indicates a further dramatic departure from the Seals test. Although both measures are visual-spatial in nature, the Balance Beam involves complex visual-spatial judgments where solution to almost half the items is contingent on balancing distance from fulcrum with the number of weights on each side. When there are unequal weights with unequal distances from the fulcrum, the complex decisions need to be based on the laws of physics. This adds a complexity to the task which is not present in the

Seals test. Wang (1987, p199) for example, maintained that a measure of frontal lobe function should have low "perceptual abstraction and eduction requirements". He defines "eduction" as a process of producing varied and pertinent solutions based on limited information. In his analysis, Wang (1987) cites the Halstead Category Test as one where the perceptual abstraction and eduction requirements are such that it would be precluded from being a measure sensitive to prefrontal lobe deficits. "When a concept formation test is highly loaded with an intelligence factor, it runs the risk of becoming an alternative form of an intelligence test rather than ... a test designed to ascertain frontal lobe function" (Wang, 1987, p199). The Balance Beam test may involve a high demand on 'eduction' processes given that subjects are required to discover the rules based on limited and specific information, and the task involves complex visual-spatial judgments.

The cognitive load demanded of the Balance Beam appears to be greater than that required for successful solution of the Seals test. Although Piaget used tasks based on the laws of physics to demonstrate Formal Operational Reasoning (Siegler, 1981), he made no mention of the complexity of these measures.

In order to gain an understanding of the complexities of the Balance Beam task, it is useful to revisit the example, presented in Section 2.5, of a child attempting to solve the Balance Beam task for the first time. Suppose that the child is then presented with the Balance Beam problem, where equal disks are placed on unequal distances from the fulcrum. To solve the problem successfully, the child is required to attend to one dimension - distance from the fulcrum. She uses her previously correct strategy of counting the disks and holds this scheme in active memory while assimilating the information before her to this existing strategy. She is wrong! The child then rethinks the possible strategies for successful solution of the task. This 'rethinking' involves an attempt at accommodating the information before her by restructuring her understanding or developing a new competency. It is argued that the 'accommodation' process of mental change involves planning, monitoring, and an active motor set. The child now searches, through active planning, to discover a new strategy. She knows from the feedback that the number of disks is not a good strategy. She can still only attend to one dimension, and soon hits on the notion of considering distance from the fulcrum. This works! She has now accommodated her strategies to suit the new information before her, and, in so doing, has learned a new strategy. The process of cognitive change involved an active search through planning, and considering the feedback from the environment to adjust behaviour.

The same child is then presented with a Balance Beam problem where unequal disks are placed on pegs which are not equi-distant from the fulcrum. She uses her previously successful strategy of counting the distance from each side of the fulcrum and discovers that this no longer works. She no longer has a strategy to solve the task. There is now a mismatch between the cognitive demands of the task, and the child's current repertoire of schemes or strategies for solving the task. The child is unable to take account of and coordinate two or more changes that occur simultaneously - number of disks and distance from the fulcrum. She is unable to "de-centre" (Wood, 1988, p41), a skill that requires attending to more than one dimension or point of view at the same time. The child's attempts at accommodating the new information are unsuccessful.

To solve the Balance Beam, at this point, requires an understanding of the laws of physics. That is, knowledge that distance from the fulcrum, as well as weight, interact to produce a solution[1].

Siegler (1981) demonstrated that even 17 year old subjects are not able to successfully solve the Balance Beam. For those subjects who were not using Siegler's rule four (see footnote), their responses on items involving unequal disks placed on pegs that were not equi-distant from the fulcrum were subject to random guessing.

To return to the example of the child struggling with the Balance Beam task, it appears that on those conflict-type items (refer to Table 1), the majority of 7-12 year old subjects would not be using the correct formula, and their responses would be trial and error guesses. Not one of the 183 subjects scored 100% accuracy on the task, attesting to the efficacy of this claim.

Support for the CN Model also comes from the prediction that the adult neuropsychological measures would not be as sensitive to detecting executive function deficits in a paediatric population. The WCST, the Austin Maze and the AVLT did not discriminate between the brain-injured and the matched normative subjects. Measures of sensitivity for these neuropsychological tests are quite low, with 0.41 for the WCST and 0.31 for the Austin Maze.

---

[1] In fact, the formula for solution is:

1. number of pegs from fulcrum x weight = torque.
2. The side with the larger torque will drop down.

### 9.3 Executive Function and Active Memory

In the context of the present study, the assertion that active memory is critical in the developmental sequence, is difficult to demonstrate. Although it is contended that those variables which load onto an Executive Function Factor on face value all had a major active memory component in addition to planning processes, it is not the intention of this study to demonstrate a double dissociation between active memory and planning. Research has been cited (Welsh, 1991), indicating that the particular presentation format for the TOH significantly affects subjects' performances. In the Welsh (1991) study, subjects are presented the TOH in ascending order of difficulty to expose children to simpler problems that are embedded in more complex subsequent problems. In contrast to previous research, **the present study**, presents the task in descending order of difficulty (Klahr and Robinson, 1981, Borys *et al* 1982). In fact, Welsh (1991) provides every opportunity for her subjects not to use active memory and claims that her presentation reflects planning processes only. She states that "an interactive testing situation often yields better test performances in young children than would be predicted" (Welsh, 1991, p71). It is suggested, however, that in the active memory mode, the TOH reflects the underlying substrates of executive functioning, in that it involves the interaction of both planning processes and active memory. The performance of subjects on the TOH in the present study lends support to the CN Model. By involving active memory it is asserted that the demands of the task are increased and its discriminatory power enhanced.

### 9.4 Seals

#### 9.4.1 Error Patterns of Normative Sample

Error patterns of the study sample were analysed using subjects' responses on the Seals task. Of those experimental developmental tests administered, the Seals task is the only one which involves a series of moves toward an end-goal, and for this reason was chosen for an examination of those particular moves. The Balance Beam, on the other hand, consists of one response for each trial presentation, and is considered to be not particularly useful for a detailed error analysis.

These error patterns, taken from the first occurring error, on the first and second trials of the 3-ball and 4-ball problem types are examined within each age group for the normative sample, and for the brain-injured sample. The pattern of errors on the Seals task appears to be descriptive of subjects' problem-solving processes and strategies. An analysis of most frequently occurring errors may provide information and an opportunity to discover the types of strategies employed by the problem-solvers. An examination of the first two trials ensures that all subjects would be included in the sample. It also provides an opportunity to inspect subsequent error patterns in order to determine whether there are particular problem-solving strategies that subjects are employing. Piaget (1976), for example, identified three stages leading towards successful solution of the 3-disk TOH problem. Stage I (5 to 7 years) subjects could solve the 2-disk problem with some difficulty but could not transfer to the 3-disk problem. Stage II subjects (7 to 9 years) could eventually solve the 3-disk problem although incurred errors along the way, and Stage III subjects (11 to 12 years) could successfully and rapidly solve the problem, generalise their solutions and recognise the importance of the first move. These three stages in the resolution of the Tower of Hanoi correspond to the three critical stages in cognitive development previously discussed. For example, 18% of the Early Concrete Operational Reasoners (7-8 years) were not able to solve the 3-ball problem within the six trials, compared to 3% of the late Concrete Operational Reasoners (9-10 years), and 0% of the Early Formal Operational Reasoners (11-12 years). On the 4-ball problem of the Seals, 59% of the Early Concrete Operational Reasoners, 56% of the Late Concrete Operational Reasoners, and 36% of the Early Formal Operational Reasoners were unable to solve the problems within the six trials. It is apparent that the Early Concrete Operational Reasoners are able to master the 2-ball problem, but not the 3 or 4-ball problem. The Late Concrete Operational Reasoners are able to master the 3 but not the 4-ball problem, and that the Early Formal Operational Reasoners are well advanced in their resolution of the 4-ball problem. These results from the current study provide support for Piaget's (1976) contention that there are three stages leading to successful solution of the Tower of Hanoi, and that the stages correspond to stages of cognitive development. To illustrate, the Early Formal Operational Reasoners are able to plan at least three moves ahead, base their decisions on taking account of all available information, including learning from previous trials, and employ reversible mental operations to develop a different strategy if the first did not work, skills which are necessary for solving the more complex 4-ball problem.

Borys, Spitz and Dorans (1982), in their study, concluded that the superior strategies by the more mature groups were related to increasing depth-of-search capacity (holding moves in active memory and planning moves ahead). Klahr and Robinson (1981) identified three strategies in the successful resolution of the TOH: subgoal selection, obstruction detection, and depth of search.

Within the study sample, the error patterns are consistent across the three age groups and the two problem types, with the first move posing the greatest difficulty. For example, although twice as many 12 year olds incur no errors on Trial 1 of the 3-ball problem, their error pattern as determined by percentage of errors, is comparable to the 7 year old subjects. On the 4-ball problem, the 7 and 8 year olds had great difficulty with the task with 27% of their errors occurring on the first move of both trials. None of the subjects successfully solved the task within the two trials. In comparison, the 11 and 12 year old subjects, 18% of whom successfully solved the task on the second trial, still incurred a significant error rate on the first move which is comparable to the youngest age group. This result is consistent with previous studies involving the TOH (Welsh, 1991).

*Post hoc* comparisons of the percentage error rates on the first move of the 4-ball problem indicate that the oldest age group in fact incurred more errors than the other age groups (52%, 54% and 69% respectively for 7-8, 9-10 and 11-12 year age groups respectively). This result may be explained in terms of the greater level of problem-solving sophistication of the oldest group. The 11 and 12 year olds would have a larger range of competing strategies to call on, and, particularly on the first move of the 4-ball problem, would have a longer sequence of planning moves to negotiate. The 4-ball problem requires a minimum of 15 moves for successful completion, while the 3-ball problem requires 7 moves. The increased complexity of the 4-ball problem demands a greater depth of search, that is, holding planned moves in active memory. Once the first move was given to the oldest age group on Trial 2, their error rates on Move #3, were considerably reduced in comparison with the younger age groups (13% with errors, compared with 31% and 25% for the Junior and Middle age groups). Borys *et al* (1982, p100) assert that overloading the cognitive system with depth of search requirements (as in the active memory mode), increases the likelihood that the older subjects when faced with the 4-ball problem with its longer sequences of moves, would revert to simpler “fall back” strategies.

The critical key error points across both problem types for all age groups were Moves #1, #3, #5 and #9. That is where the majority of errors occurred. It appears that despite the obvious high success rate by the older age group, the critical choice points occur at the same move positions as it did in the younger problem-solvers.

What is the relevance of these key error points to understanding the development of cognitive function? Are there changes in the cognitive demands of the tasks at these critical points which impact on changes in cognitive capacity? Are they key-error points, or critical decision points in the task, and, if so, are they significant windows into understanding the development of executive functioning?

For Move #1 in both the 3 and 4-ball problems, there are two legal moves into two empty Seals, as well as the possibility of illegal moves. Subjects are faced with a choice between two legal moves, both of which would not infringe the stated rule that "a larger ball cannot be placed on top of a smaller ball". Subjects can decide between a chance choice of 50/50, or the employment of decision-based strategies which involve "subgoal ordering, obstructor detection and depth of search" (Klahr and Robinson, 1981, p 133). These three strategies involve planning. All of the high error choice points contain significant shifts in cognitive demand where subjects must employ planning processes involving specific strategies.

For example, Move #3 in both problem-types involves placing the smaller green ball back onto the medium-sized red ball to free a space for the larger blue ball. Subjects must employ an "obstructor detection and removal" strategy (Klahr and Robinson, 1981, p133) in order to recognise that the green ball needs to be removed from its position, and to recognise that the blue ball needs to be cleared so that it can consequently be placed on the third Seal. This is a complex decision in the 3-ball problem as it involves moving a ball **backwards**. Up to this point, all moves are in a forward direction. A shift in cognitive set is required at this point to be effective. As well as involving this lateral thinking, Move #3 also presents the subject with two possible legal moves. Move #3 signals the need to employ a new cognitive strategy. Following Move #1, it is contended that Move #3 is the next critical choice point where subjects must employ a different type of strategy for solution in the minimum number of moves.

Move #5 for both the 3 and 4-ball problem also involves an obstructor detection and removal strategy. In the 3-ball problem this involves clearing the green ball into an empty seal so that the red ball can be placed on top of the larger blue ball. In the 4-ball problem this is a critical move as it involves clearing the green ball so that a pyramid subgoal can be built to free a space for the largest yellow ball to be moved to the third seal. The subject must think in terms of moving a sequence of balls or in Anzai and Simon's (1979) terminology "moving pyramids".

Move #9 in the 4-ball problem involves what may appear to be a blocking move by placing the green ball on top of the yellow ball rather than onto an empty Seal. This move enables the red ball to be placed on the empty Seal so that the blue can be moved on top of the yellow. It involves the planning of intermediate goals in a means-end solution and requires the subject to think ahead several moves. These later moves take on greater complexity in this active memory presentation as subjects need to hold the moves in active memory, referring back to previous moves, while planning their next moves. In summary, it is proposed that moves #1, #3, #5 and #9 represent decision points where subjects are faced with two legal moves. The response must be based on complex sequential steps where these particular moves are sub-goals or preliminary steps toward the final goal-state. Subjects may use a variety of strategies, consistent with the stated rules of the task. It is contended that these strategies involve the interplay between active memory and planning processes.

Observations of a subject performing the TOH in other research identified that all additional talk, lengthy pauses and negative comments occurred on moves 1, 5 , 9 ..4k + 1 (Anzai and Simon, 1979, VanLehn, 1991). VanLehn (1991) reanalysed the Anzai and Simon (1979) data of a child learning the Tower of Hanoi and discovered a pattern which cut across all phases of the TOH. He used the analysis of one subject's protocol over a 90 minute period and found that all the subject's additional talk, lengthy delays and negative comments occurred on moves 1, 5, 9, ..4k+1. VanLehn's (1991) proposal on the importance of the 4k + 1 moves in solving the Tower of Hanoi was previously discussed in Section 2.5.1. VanLehn's subject appeared to struggle at each of these move points. At each of the 4k+1 moves, the subject has a choice between two legal moves (in addition to illegal moves). A correct decision is critical to the overall successful solution to the task, although a wrong choice can still result in finishing the task but not in the minimum number of moves.

VanLehn speculated that correct solution at move 4k+1 involved strategy modification where the subject "interrupts her normal problem solving, switches her attention to the problem of modifying her strategy and begins to mutter goals appropriate to that task instead of her normal disk-moving goals" (p5). VanLehn's subject employed a “scientific discovery” approach during these moves and set up experiments to test her reasoning. This ‘discovery’ behaviour is comparable to hypothetico-deductive reasoning, a ‘hallmark’ of Formal Operational Reasoning (Ginsburg and Opper, 1988).

It might be expected that the active memory component of the Seals task employed in the current study differs from the Anzai and Simon presentation. In the current study, greater stress would be placed on the key error points when active memory is employed. This hypothesis may account for the result that Move #3 is also identified as a key choice point critical to successful solution in the current study as subjects have to hold a number of moves in active memory.

If the argument is accepted that strategy development occurs at the critical decision points in solving the Seals task, then how does this notion fit with the current results (and those of others, such as Welsh, 1991), that the majority of error moves was invariant across the three age groups? An examination of Tables 29 and 30 highlights the percentage of errors that was made for each age group in solving the 4-ball Seals task. Data for Trials 1 and 2 are presented. The first comment to be made from a comparison of both tables is that all three age groups appear to struggle at the critical error points. In fact, a higher percentage of the 11-12 year olds incurred errors on Move #1 than did the other two age groups (69% errors for 11-12 year olds, 54% for 9-10 year olds, and 52% of errors for 7-8 year olds). On subsequent moves during Trial 1, the percentage of errors for the 11-12 year olds is less than for the other age groups. Similarly, on Trial 2, presentation of the 4-ball Seals task, the percentage of 11-12 year olds with errors on Move #2 and #3 is less than the other age groups, and then on Move #5, marginally more 11-12 year olds incurred errors than did the other age groups. The current error pattern can be explained in terms of the proposed CN Model.

The 11-12 year olds, who are at an early Formal Operational Reasoning period of cognitive development, are faced with a task of planning a fifteen move problem. They not only have to hold the current moves in active memory, but use reversible mental operations to remember their previous moves, so that they can experiment with a different first move if the current one fails. Planning for the 11-12 year old involves taking all the possible moves into consideration, weighing each against the stated rules of the task, and considering the likely outcome. This is the ‘scientific discovery’ observed by VanLehn (1991) when his subject was solving the TOH. It is speculated that the underlying neural processes of premotor and dorsolateral prefrontal cortex interact in a continuous feedback operation to complete the “Perception-action” cycle so that active memory and planning processes can occur to solve the task.

Maturational factors serve to increase active memory and planning capacity, and therefore provide an increased range of cognitive strategies to deal with complex cognitive demands. In the active memory mode, the TOH challenges the child not only to plan but to hold in active memory previous as well as present moves. It is suggested that, compared to the 11-12 year old who would be applying ‘scientific discovery’ to the task, the 7-8 year old would be employing a ‘trial-and-error’ approach. Given that there are two possible correct moves on Move #1, the younger age group has a 50/50 chance of success and it is conceivable that their guesses on this first move would be more successful than the 11-12 year old’s struggle to logically derive the correct response. If this logic is accepted, then it would explain why more 11-12 year olds incurred errors on Move #1 than did the other two age groups.

There is another explanation for the apparent difficulties experienced by the 11-12 year olds, when compared to the other age groups. Up to a certain point in the Seals task, it is argued that, with increasing age, there is an apparent inverse relationship between success and number of moves to resolution. This phenomenon may explain the difference in performance between the three age groups. Because the task is presented in descending order of difficulty, and there have been four critical choice points identified in the 4-ball problem (Moves #1, 3, 5, and 9), then it is asserted that, if subjects can successfully solve the problem up to and including Move #5, there is only one more critical choice point to negotiate - Move #9. It is contended that Move #5 represents the 'turning point' when the cognitive demands are within the capabilities of the 11-12 year old Early Formal Operational Reasoner. From Moves #1 to 5, it is contended, the task changes from one which is cognitively challenging to the 11-12 year olds, to one which is presumably within their capacity. It follows, that from Move #9, the 4-ball problem would be an easier task to solve as subsequent moves do not place as heavier a tax on planning processes as those earlier moves. Consequently, the graphs portrayed in Figures 51 to 53 reveal a rise in the percentage of errors up to Move #5 with a decline on Move #9. The assertion that the difficulty level of the 4-ball Seals task increases up to Move #5 and then declines, however, was not empirically tested, and is possibly a study for future research. Figure 51 portrays the success rate for the 11-12 year olds on the 4-ball problem.

From Figure 51, it is apparent that the difficulty level of the 4-ball Seal task for the 11-12 year olds rises at each of the critical choice points and peaks at Move #5. This pattern is similar to that of the 9-10 year old group, but quite different to that of the 7-8 year group. Figures 52 and 53 represent the pattern of errors on the same task for the younger age groups.

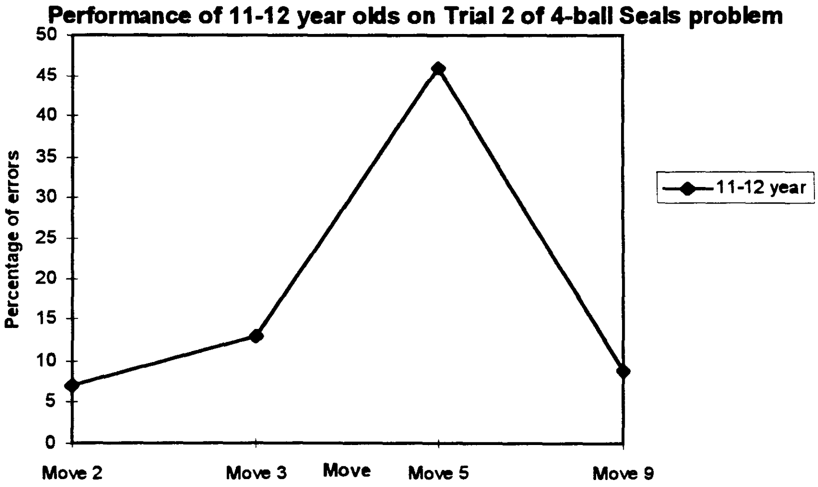

*Figure 51: Percentage of 11-12 year olds scoring errors on moves of the 4-ball Seals task*

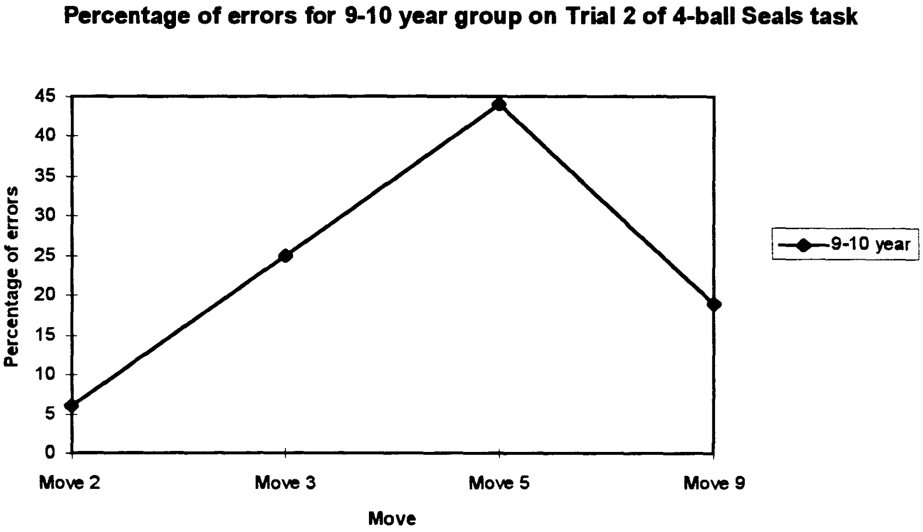

*Figure 52: Percentage of errors for the 9-10 year group on Trial 2 of the 4-ball task*

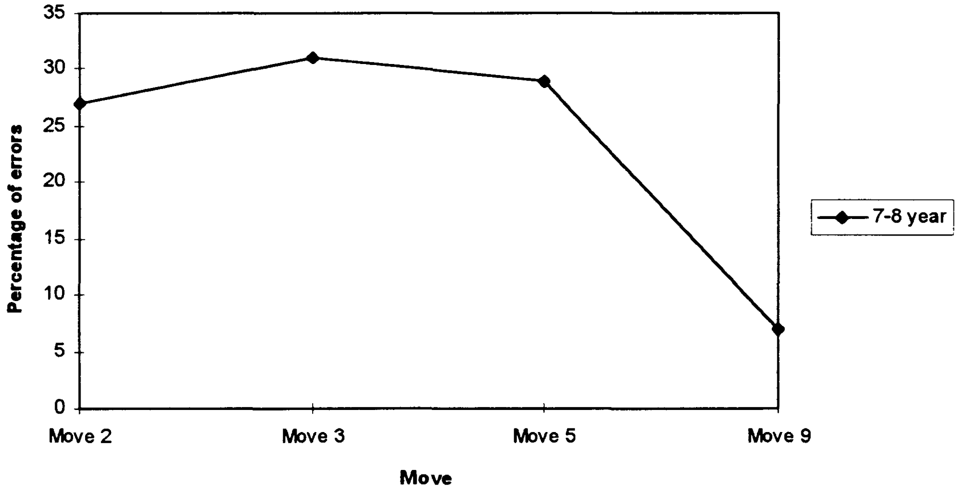

*Figure 53: Percentage of errors of 7-8 year olds on Trial 2 of 4-ball Seals task*

Figure 53 is a little misleading in that the reported drop in errors on Move #9 is an artifact, caused by the fact that the majority of 7-8 year olds would not have reached Move #9 by the six trials for the 4-ball problem. Their error scores would therefore not have been included in the result. The youngest age group incurred approximately the same percentage of errors up to Move #5, suggestng that they are probably responding either to a trial-and-error chance strategy, or a learned strategy that was ineffective in the situation (for example, "move the ball to the nearest free space"). There is therefore, little evidence of learning over the first five trials. It is contended that this result is consistent with Piagetian cognitive development in that the child at an Early Concrete Operational Reasoning period would consider each move as quite separate from preceding ones, and make a judgment based on the information in front of him. The 7-8 year old's poorly developed capacity for active memory and planning, in comparison to the older age groups, would compromise his performance on the 4-ball task. Figure 53 is a little misleading in that the reported drop in errors on Move #9 is an artifact, caused by the fact that the majority of 7-8 year old subjects had already reached a ceiling by Move #9, and their error scores would not have been included in the result.

In summary, the results of error analysis for the three age groups on the 4-ball Seals task suggest that although the percentage of errors on the critical choice points appear to be invariant across the age groups, there is evidence that a difference in strategy application between the youngest and oldest age groups does exist. In particular, the profile of the 7-8 year olds is quite different to the older two age groups, and does suggest that a difference exists between the type of strategies employed by the youngest age group. Further research would be required to specifically tease out the different strategies employed by the three age groups at the critical choice points in the resolution of the Seals task.

### 9.4.2 Self-Corrections

An analysis of the self-corrections made by all subjects during the first two trials of the 3 and 4-ball problems of the Seals task is enlightening in that self-corrections occurred only at the key error points. Only the first self-correction for each subject was included in the analysis and, as self-corrections appear to be evenly distributed across both problem types, the results were aggregated for each age but not across ages.

For each age group, self-corrections appear only at the key error points of Moves #1, #3 and #9. The incidence of self-corrections appears stable for the youngest and middle age groups, and rises for the oldest age group with 11-12 year olds having more than double the number of self-corrections than 7 year olds. This may reflect a greater self-monitoring ability for the older subjects.

It is also at these moves, where the majority of verbalisation occur, although this is a weak statistic because internal self-talk cannot be monitored directly. 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 Subject #218 then moved correctly. Subject #215, on Trial 2 of the 3-ball problem: "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.

VanLehn (1991) described the critical decision points as an *impasse,* which is defined as where "the cognitive architecture cannot decide what to do next given the knowledge and situation that are its current focus of attention." (VanLehn, 1991, p19). He cites evidence that subject's verbalisations and lengthy delays in responding only occur at these critical choice points. Evidence is also presented that strategy acquisition, that is, the learning of a new strategy for solving the TOH, occurs at Moves 4k +1.

VanLehn's hypothesis can be extended with an explanation in terms of Piagetian theory. The 4k+1 moves appear to represent critical choice points in the resolution of the TOH. As VanLehn proposes, it is here that subjects use hypothetico-deductive reasoning to experiment with particular strategies.

The two strategies which involve planning identified by Borys *et al* (1982) are examples. If the 4k+1 points represent an *impasse*, then they create 'disequilibrium' within the subject. That is, there is internal conflict created by a mismatch between the cognitive demands of the task and the subject's current store of strategies for solving the particular task. This internal conflict is a signal for the interaction of accommodation and assimilation processes to move the subject through this 'disequilibrium' so that mental growth can occur. As the action component of cognitive growth, accommodation processes, through planning and experimentation, construct new mental structures (in this case new strategies or rules for solving the TOH) to move the subject through the *impasse*. The administration of the TOH in the active memory mode would have the effect of placing greater stress on the 4k+1 moves, particularly in the 3 and 4-disk problems, as subjects have to hold more moves in active memory while applying planning processes. It is hypothesised that when the Tower of Hanoi is employed in the active memory mode, the 4k+1 moves are a window into the development of cognitive growth. Consequently, an analysis of subject's moves on the critical choice points of the Seals task provides important information on the type of strategies employed, which in turn leads to an understanding of the subject's cognitive stage of development. The error patterns on these 4k +1 moves provide information on the types of strategies that the subject is using (rule violations, subgoal selection, obstructor detection and depth of search).

It is also contended that the "cognitive architecture" for explaining impasse-driven strategy acquisition described by VanLehn (1991) can be explained in terms of the CN Model. In the case of the child in the VanLehn study trying to resolve the TOH puzzle, the CN Model presented in the current study posits that information on the task is gained through sensory channels from the environment and then processed through two independent but interrelated processors. The neurological systems are premotor and dorsolateral processors, where information is passed backwards and forwards in an interactive loop. The parallel psychological systems are accommodation and assimilation processors which have cognitive functions of active memory and planning. As the child interacts with the TOH task, previously learned strategies come into play to resolve the problems. However, as the cognitive demands of the task become more complex, the child reaches an *impasse* where previously learned strategies are no longer effective in solving the task. There is now a mismatch between the cognitive demands of the task and the child's store of learned strategies. This is the point of 'disequilibrium', described by Piaget (1976) as internal conflict created by an inability to accommodate new information to existing mental structures. This is also the point of '*impasse*' described previously by VanLehn (1991). It is at this point that the CN Model demonstrates its dynamic nature. The point of conflict triggers the interaction between assimilation and accommodation processors (with correlates in premotor and dorsolateral prefrontal cortex processors) to move the child to experiment with and to develop new strategies in order to overcome the internal conflict. This experimentation involves the interaction between active memory and planning, coordinated by executive functioning, to produce cognitive growth and move the child through to the next cognitive stage of development. It is acknowledged that this may be an oversimplification of the process which in reality takes many months to move the child to a higher stage of mental growth. As Siegler (1991, p45) comments, the child does not suddenly "wake up out of the blue with two new learned strategies!"

## 9.5 Relationship of Executive Function to School Ability and Memory

There is a strong relationship between the Executive Function factor and school ability and memory. This relationship increases with age. It involves a significant and positive linear relationship with both variables. All the variables which load significantly on the Executive Function factor have a positive relationship with the OLSAT Total test score. Regression analyses confirm a significant linear relationship between OLSAT and Balance Beam and OLSAT and Seals. In addition, Pearson correlation coefficients, within each of the three age groups, display an increasing and positive correlation between OLSAT and Balance Beam scores. Correlations are $r=0.355$, $r=0.434$ and $r=0.641$ for the Junior, Middle and Senior age groups respectively. A Fisher Transformation R to Z confirmed the significant relationship between the Junior and Senior age groups. Two other variables which load onto Executive Function, Austin Maze mean errors and Austin Maze time taken, also had positive and significant Pearson correlation coefficients with OLSAT Total scores. Maze mean errors per trial display significantly different correlations between the Junior and Senior and the Junior and Middle age groups, with increasing correlations with increasing age. This result is also confirmed by the Fisher Transformation R to Z technique and establishes that executive functioning plays an increasingly important role in the successful performance of school-related tasks. As the cognitive demands of academic tasks increase with age, so too does the importance of executive functions to the successful solution of those tasks. This result suggests that executive function and school ability are not independent variables, as postulated in the second hypothesis of the present study. There is clearly a positive and increasing developmental relationship between the two.

The relationship between Executive Function and school ability, as measured by the OLSAT, indicates that OLSAT scores can be positively predicted from scores on Executive Function variables. Competence on executive functioning is positively and significantly related to school ability. This result appears at variance with previously reported studies (Williams and Mateer, 1992), where deficits in executive functioning were not associated with concomitant deficits in academic performance. In the Williams and Mateer (1992) single case studies, both students maintained superior levels of academic and intellectual performance while disruption to executive functioning resulted in poor adaptive behaviour. However, the Williams and Mateer study involves a number of significant departures from the present study. It examines two case studies of children who sustained traumatic brain injury, as opposed to the present analysis of a non-clinical sample of children. In addition the authors' description of executive function is neither theoretically nor neurologically based, and measures of executive function are based on a loose set of criteria involving observation, behavioural rating scales and adult neuropsychological tests. In addition, the tests that are used in the current study are different to those used by Williams and Mateer and consequently may have tapped into different variables. Caution should therefore be exercised in making a direct comparison between the results of the present study and those of Williams and Mateer (1992).

The positive relationship between school ability and executive function is further supported by results at an individual test level. There is evidence that Pearson correlation coefficients between OLSAT Total score and Balance Beam scores represent an increasingly statistically significant relationship as you move from the younger age group to the older normative group. Fisher R to Z transformations confirm that the correlations of $r = 0.355$ and $r = 0.622$ for the Junior and Senior age groups are statistically different at the $p<0.05$ confidence interval. This suggests that performance on the Balance Beam is increasingly related to school ability (as measured by OLSAT scores), as the ages move from 7 years to 12 years. A caveat should be made at this point. The result **may** also be an artifact of the different OLSAT tests administered to the two age groups. Details of the different test forms are presented in Section 4.1.8. The Senior OLSAT test contains more problem-solving items than are contained at the Junior level. For example, eighteen Quantitative Reasoning items and twenty Verbal Reasoning items are contained in the Senior version of OLSAT, while none are included in the Junior version. On face value, performance on both OLSAT subtests would be expected to correlate with the Balance Beam test, and may have contributed to the positive correlation. Consequently, as the OLSAT items are not held constant across the age ranges, it is difficult to make a valid comparison for age differences in OLSAT performance.

Pearson correlation coefficients between OLSAT Total score and the Seals task represent a consistent relationship across the three age groups with correlations of r = 0.383, r= 0.325 and r= 0.359, for Junior, Middle and Senior age groups. This result also provides some support, at an individual test level, for a positive relationship between school performance and ability on a measure of experimental developmental function. In summary, there is supporting evidence that developmentally, school-related ability is positively enhanced by executive functioning. Successful performance on school-related activities places an increasing demand on both active memory and planning processes as the tasks demand higher levels of both components with increasing age.

The relationship between Executive Function and memory was also examined using correlational and regression analyses. Results indicate a significant and positive relationship between the underlying Memory/Learning Factor and the Executive Function Factor. Correlations between the two factors for each of the three age ranges are reported at r= 0.309, r= 0.457 and r= 0.596 for the Junior, Middle and Senior age groups respectively. Fisher R to Z transformations indicate a significant difference in correlation coefficients between the Junior and Senior age groups. To confirm this relationship using individual variables which load significantly onto the Memory Factor, regression analyses produce a significant predictive value for both Seals and Balance Beam for memory performance as measured by the AVLT Learning Index. Significant regression equations predicting memory scores from both Seals and Balance Beam indices are found for the Junior and Senior age groups. A somewhat unexpected result is the comparatively low Pearson correlational coefficient (r=0.031) between Seals and AVLT for the Middle age group. Moderate correlations of r=0.395 and r=0.272 were computed for the Junior and Senior age groups respectively. It appears that the 9-10 year old subjects are functioning quite differently to the other age groups. This point is taken up in Section 9.11.

The results of the relationship between executive function and memory, taken from individual variable correlations, underlying factor scores and regression analyses, indicate that competence on memory tasks is positively related to ability to perform on measures of executive function. Given the underlying theoretical model that executive function involves the interaction between active memory and planning processes (Fuster, 1993), this result is viewed as providing support for the CN Model. From the model proposed in the current study, it is anticipated that a positive correlation would exist between executive function and memory, and, as both psychological constructs are developmental in nature, this relationship would increase with age up to adolescence. *Post hoc* comparisons of the correlations between Executive Function and Memory Factors indicate an increasing correlation coefficient for age for each of the three age groups. However, only the Junior and Senior age groups are significantly different.

It is suggested that the underlying neural mechanism for the development of memory becoming increasing related to the development of executive functioning, is the myelination of nerve fibres in frontal connections. For example, there is evidence that the timetable for myelin development may account for the increase in speed and efficiency of responding as well as increase in memory capacity (Gibson, 1991). In the human, myelination occurs from the age of 2 years to adolescence, beginning in the primary sensory and motor areas and spreading to the association areas (Straudt *et al*, 1993).

## 9.6 FISH TASK

The results of the Fish task indicate a developmental improvement in scores with increasing age, although statistical significance, at the Bonferroni adjusted level, is demonstrated only between the 9-10 and 11-12, and between the 7-8 and 11-12 year age groups. It appears that there is a significant rise in performance from 9-10 years of age. There is also a statistically significant improvement in Fish Perseverative Error performance from 9-10 years of age. See Figures 15 and 16 for a display of Fish results.

As a test, the Fish task has good internal reliability with a Cronbach alpha internal reliability coefficient of 0.886. The Rasch analysis (Rasch, 1966) provides a goodness of fit for each item to a unidimensional model. The Rasch process places each item on the same linear scale so that a measure of item difficulty, across the 183 subjects, can be obtained. Large negative scores (>-3) indicate that the items have excessively orderly responses and large positive Rasch scores (>3) indicate a large variability in responses where high scoring subjects incurred low scores and low scoring subjects incurred high scores. Eleven of the thirty-five items have Rasch item difficulty scores >-3 indicating they have a poor goodness of fit to the Fish unidimensional model. These eleven items are measuring something quite different to those which have Rasch scores within 2 standard deviations from the mean, and consequently, have poor discriminatory power.

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

A possible explanation for this outcome is that subjects inhibit responses to the Background fish by not clicking on the mouse, and when a Target fish immediately follows are already in response inhibition mode. They presumably hold back their response until they receive 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 have just clicked on a Target fish, are in "go" mode and continue to click on presentation of a Background fish. Reaction time to each stimulus would have confirmed this hypothesis. Unfortunately, it was not measured in the current study.

It is considered that, given the good internal reliability of the Fish task (r=0.886), the result that it loads significantly on a separate Inhibition factor (r=0.717), and that it provides a separation between the three age groups, the Fish task is one which is worthy of modification for future research. Given that the Fish task is developmental in nature and that it clearly loads onto a separate Response Inhibition factor, it is considered a measure which appears to examine the development of response inhibition in a group of 7 to 12 year old males. The underlying factor loadings, set out in Table 60, clearly indicate that the Fish task has negligible loadings on either the Memory/Learning factor or the Executive Function factor, with loadings of 0.177 and -0.070, respectively.

It is suggested that the Fish task taps one component of the planning process which is identified in the CN Model. Fuster's (1993) planning process of active motor components of behaviour includes monitoring of behaviour, adjustment of responses in light of environmental feedback and response inhibition. It is contended that response inhibition is therefore but one aspect of the planning process and that the Fish task has clinical utility as a developmental measure of response inhibition.

It is proposed that the eleven identified items from the Rasch analysis be targeted for modification to the Fish task. As previously discussed, items which are excessively orderly in their responses, are those Target fish which are immediately followed by a Background fish. It is proposed to redesign the task by rearranging the Background fish which occur prior to the eleven poorly fitting items, so that more Target fish are placed in a consecutive string.

## 9.7 PIGGY BANK

The results of the Piggy Bank Task suggest an incremental improvement in scores with increasing age. This developmental progression is statistically significant at the Bonferroni adjusted level, between the 7-8 and 11-12 year age groups, but not between the 9-10 and 11-12 and between the 7-8 and 9-10 year age groups. Figures 17 and 18 detail the Piggy Bank results for the three age groups. Variance does not explain the lack of developmental improvement in scores for the 9-10 year old group. The Piggy Bank task is one of the few measures where the 9-10 year old group display standard deviation scores which decrease with increasing age.

The test has a low internal reliability coefficient of 0.392, indicating that it is unstable. Individual item Rasch analysis scores indicate that four of the ten items, #2, #3, #5 and #6 had low negative fit scores (>-3), suggesting that they are poor discriminators. An inspection of these particular trials reveals that three of the four (#2, #5 and #6) involve no movement of the piggy bank with the coin. That is, the piggy bank which contains the coin did not move from its initial position. These three trials are the only ones which involve no movement. It has been argued that trials involving no movement (there were still two transpositions of piggy banks) are treated as an additional distractor to remembering which piggy bank had the coin. It now appears that these trials have poor discriminatory power. In addition, the Rasch statistics indicate that items #7 and #8 are the two which are the closest to the predicted unidimensional model, with Rasch scores of -1.953 and -1.588, suggesting they are good predictors of the model (Wright and Masters, 1982). See Table 18 for full details of Rasch scores. Both items involve a twenty second delay between final transposition and reappearance of the five piggy banks, both involve one movement of the coin, and in both cases, the transposition of the coin occurs on the first movement. Position of the coin into the piggy bank did not appear to be an issue as item #7 involves the coin placed into the last piggy bank and item #8 involves the coin placed into the middle piggy bank. It is also noted that in addition to having good discriminatory power, item #8 also scores the highest internal reliability coefficient of 0.44 (compared to 0.392 for the overall test).

There is some question whether the Piggy Bank task has adequate internal test construction properties, in that it has a low internal reliability coefficient (Cronbach alpha = 0.392). It is contended, however, that it should be modified, taking into account the points already discussed. Items #2, #3, #5, and #6 would be replaced with new items. It is proposed that each of the four new items would involve one movement of the coin on the first transposition, and the five piggy banks would reappear after a 20 second delay period. These criteria are chosen from the results of the Rasch analysis which indicates that items which appear to conform to the Piggy Bank unidimensional model are those which involve a 20 second delay, involve one transposition of the piggy bank with the coin, and that this transposition occurs on the first movement. Presumably, there is a greater memory demand placed on the first transposition, as subjects are then presented with the distracting second transposition which does not contain the piggy bank with the coin. They have to retain the memory of the position of the coin during this distraction, and then during the 20 second delay.

The results of the underlying factor structure, detailed in Table 60, indicate that the Piggy Bank task loads significantly on a Response Inhibition factor, with loadings approaching zero on the Memory/Learning and Executive Function factors. As with the Fish task, the Piggy Bank task appears to be a developmental measure of response inhibition. Within the proposed CN Model, the Piggy Bank task taps into one component of the active motor set planning process.

## 9.8 BALANCE BEAM

The results of the Balance Beam task demonstrate its sensitivity to developmental changes. Scores are statistically significant, at the Bonferroni adjusted level, between the 7-8 and 9-10, and the 7-8 and 11-12 year age groups. The difference in scores between the 9-10 and 11-12 year groups approaches significance (p=0.049). The Balance Beam task is considered to have good discriminatory properties between the three age groups of healthy subjects. See Table 20 and Figures 19 and 21 for details of results.

Item difficulty levels using Rasch analysis were developed across all subjects. 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 are >0.920. The mean score for this group of twelve trials is 0.960.

Two points arise from this result:

The first eight trials are all of the Dominant type with solution based on number of disks. Recall that the Dominant problem type occurs when unequal disks are 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 are of the Subordinate type. Subordinate problem types involve equal disks which are placed on non-corresponding pegs. Solution is also one dimension- distance from the fulcrum. In these four trials, disks are equal on both sides of the fulcrum although placed on noncorresponding pegs.

Successful solution to the first eight trials of the Balance Beam is based on an examination of the number of disks on each side of the fulcrum. Subjects need only be aware of responding along one dimension. However, if they continue 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 is incorrect. This is because Trials 9 to 12 now require subjects to change their judgments from the dimension ‘number of disks’ to ‘distance from fulcrum’. Correct solution is still unidimensional, as the number of disks remained constant, although the dimension has changed. This switch in category may account for the low score on Trial 9.

The second difficulty stage involves the remaining twelve trials with the exception of Trial 18. Rasch difficulty scores for this bracket of trials are within 2 standard deviations above and below the mean (range of 2.179 to -1.709) suggesting that these eleven items are discriminating well and are 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. Subjects must now take account of two dimensions.

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 is 5.404 suggesting that this trial does not conform to the other trials. Mean score across all subjects is 0.005 suggesting that very few subjects were correct for this trial.

Overall, the Balance Beam has moderate internal consistency with a Cronbach alpha coefficient score of 0.561. The Rasch item difficulty scores indicate a two stage difficulty level for the test. It is contended that the coefficient alpha score may have been affected by the two difficulty levels. To partition this effect, internal reliability coefficients were computed for each difficulty level of items, across the 183 subjects. Reliability coefficient for the first level (Items 1-12) is 0.635 and a reliability coefficient for the second difficulty level (Items 13-24) is 0.445.

Although the Balance Beam task loads significantly onto the Executive Function factor, with a loading of of 0.526, as outlined in Table 60, it also has a loading of 0.410 on the underlying Memory/Learning factor. This result would suggest that the task taps into both active memory/planning processes as well as memory/learning. It was suggested, in Section 9.2, that the Balance Beam task may contain a number of complexities which would preclude it from being a strong measure of executive function. From the results of the present study, it appears that the task may be more eductive than hypothetico-deductive, a point which would make it incompatible with the CN Model. The Balance Beam task involves complex laws of physics for successful solution and subjects are required to discover the rules of the task rather than applying a given set of rules. According to Wang (1987), tests which are eductive tend to be alternative forms of intelligence tests. However, the relationship between performance on the Balance Beam and a measure of intelligence was not formally assessed. It is suggested that this is a point which should be taken up in further research.

## 9.9 AVLT

The results of testing on the AVLT provide support for developmental improvement in two indices: Learning and Retrieval. In the present study, significant improvement in scores did not occur until 9-10 years of age on both indicators, suggesting that there is a two-stage level of maturation in performance on these measures. No significant difference is demonstrated in scores between the 7-8 and 9-10 year old age groups, although a *post hoc* inspection indicates a trend toward increased improvement in both cases. The first stage of maturation therefore involves subjects in the 7-10 year old age level, and the second maturational stage involves subjects in the 11-12 year old age level. It can be concluded from the present study, that the AVLT is not sensitive to developmental changes in subjects below the ages of 9-10 years. Geffen and her colleagues found that both Learning and Retrieval indices of the AVLT are sensitive to age improvements from 7-12 years of age, a result which is at variance with the present study. The difference between the current study and that of Geffen *et al*, (1993) is identified in the 9-10 year age range, with a mean of 39.6 words remembered for the present study versus 47.3 for Geffen *et al* (1993) for the Learning variable, a difference of one standard deviation. Similarly, in the Retrieval category, the Geffen *at al* (1993) norms are approximately 1 standard deviation higher than in the present study for the 9-10 year old group (0.57 for present study versus 0.71 for Geffen). An inspection of the remaining AVLT indices indicates that the two samples are remarkably similar in their performance results. So the difference between the present 9-10 year old group and Geffen's 9-10 year old boys are specific to the two indices which display sensitivity to age effects.

It is contended that the significant variability in performance of the 9-10 year old subjects may have accounted for the lack of significant difference between this age group and the other two groups. This point will be taken up in Section 9.11. It is also contended that the relatively small sample size of the Geffen *et al* study, with only ten males in each age category, compared to at least sixty in the current study, may have contributed to the difference in results. In addition, the sample in the present study was carefully chosen to represent children drawn from a population that is considered representative of the Sydney City population, as determined by the Australian Bureau of Census (1991). On the other hand, the sampling basis in the Geffen study is not clear.

What do the results of the AVLT mean in terms of the CN Model? From the predictions made in Table 2, it was asserted that the AVLT would not be a measure of executive function in a paediatric population. Table 60 confirms that the AVLT loads significantly onto an underlying Memory/Learning factor (r =0.619), although not on an Executive Function factor (r =0.294). In this regard, the predictions from the CN Model are confirmed. In addition, there were no significant differences in performance across the three age groups of subjects on any of the AVLT measures employed in the current study. It is concluded that the AVLT does not measure executive functioning in a population of children in the 7 to 12 year age range.

## 9.10 Austin Maze

The results of the Austin Maze indicate that two measures produced significant changes with age. Mean Errors per trial and Total Errors both discriminated significantly between the 9-10 year and 11-12 year old groups, and between the 7-8 and 11-12 year age groups, but not between the 7-8 and 9-10 year old age groups. This result suggests that performance on the Austin Maze is characterised by a two-stage level of maturation. The first level of performance contains subjects in the 7-10 year old age group, and the second stage of maturation contains the 11-12 year old group of subjects. The Maze is not sensitive to developmental changes in performance for subjects below the age of 9-10 years of age. However, although the difference in Austin Maze Error scores is not statistically significant for all three age groups, the general trend indicates a decrease in errors with increasing age.

An analysis of individual error types within the Austin Maze indicates 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: Not going 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.

What can we make of the results of the Austin Maze? In terms of the CN Model proposed in the current study, it is contended that there is an increased capacity, with age, to retain previous moves in active memory. This would result in less perseverative errors. Increased active memory capacity would result in a decrease in errors and an ability to solve the Austin Maze in less trials, two results which are reflected in the current study. In addition, a *post hoc* analysis of results for the Time Taken variable indicates a developmental improvement with increasing age. Mean scores for Time Taken are 1064.15 seconds, 851.36 seconds and 567.18 seconds for the 7-8, 9-10 and 11-12 year age groups respectively. A reduction in time taken to complete the task would be indicative of increased speed of information processing and would lend weight to the argument that active memory plays a significant role in the successful performance of the Austin Maze. Mean Errors/Trial and Time Taken are two Austin Maze variables which load onto the separate Executive Function factor. It has previously been established that there is a significant developmental improvement with age on the Executive Function factor.

The result that the two Austin Maze variables load onto the Executive Function factor suggests that both tap into active memory and planning processes. In order to successfully solve the Austin Maze, the child needs to hold a number of moves in active memory while planning the remainder of moves. Once the child has successfully reached the end square, subsequent trials on the task become one of encoding and retrieving the previously correct sequence using active memory. In Piagetian terms, the process of 'reversible mental operations' is paramount to retracing one's step following the end of a trial. Without this capacity to return to the beginning of the sequence and retrace one's first step, the child will be forced to use trial-and-error procedures to solve the Maze on subsequent trials. It follows that specific components of the Austin Maze display sensitivity to maturational factors in the development of executive functions. The Early Concrete Operational Reasoner, who has not developed a set of 'reversible mental operations', would find difficulty in coping with each new trial. To this child, each trial is treated as a new problem as information is treated singly, without regard to previously acquired information on the position of the correct pathway. The 7-8 year old Early Concrete Operational Reasoner would be bound by the information that is front of him or her. There would be little, if any, attempt by the Early Concrete Operational Reasoner to apply a 'scientific discovery' approach to learning the maze, in the same manner that could be applied to solving the Seals task. It is contended that the 7-8 year old would have difficulty holding previously learned moves in memory. Responses would tend to be of a trial-and-error nature. The Early Formal Operational Reasoner, on the other hand, would apply hypothetico-deductive reasoning to perceive the relationship between individual trials, would use active memory processes to hold previously learned moves in memory during the delay period between the presentation of each trial, and, at the same time, would plan prospective moves. The developmental nature of this process is reflected in significantly less errors for the 11-12 year old subjects, compared to the other two age groups.

An analysis of the 9-10 and 11-12 year old subjects' error patterns confirm that scores on the first ten trials are a good predictor of total errors over the twenty trials or when criterion was reached (Pearson correlation coefficients between scores on 10 trials and 20 trials are 0.955 and 0.956 for the two older age groups). On the other hand, scores for the 7-8 year old group are less predictable with a correlation of 0.775. Number of trials needed to reach criterion for each age group display improvement with increasing age, with a sharp reduction in trials between 9-10 and 11-12 years (mean trials were 17.9, 17.1, and 13.9 with increasing age). The performance of the 11-12 year olds is comparable with adult populations employed in the Bowden, Dumendzic, Clifford, Hopper, Tucker and Kinsella, (1992) study (mean trials of 14.2), and research cited by Walsh (1985).

Recent studies have exposed the utility of using the 'error utilisation' criterion on the Austin Maze for clinical interpretation. Normal subjects may often take 20-30 trials and occasionally up to 40 trials to obtain a criterion of two consecutive errorless trials (Bowden *et al,* 1992). In a reanalysis of Austin Maze scores of six subject groups (normal, heterogeneous neurological, alcohol-dependent, frontal lesions, closed head injury and temporal lobe epilepsy), Bowden and Smith (1994) found that after ten learning trials they were able to predict with a high degree of accuracy the number of errors incurred to criterion. This appeared true for all six subject groups. The authors concluded: "therefore, it would not seem advisable to continue localising deficits to the frontal lobes, or attributing impairment to frontal dysfunction, on the basis of a failure to observe quick acquisition or a stable error-free performance on the Austin Maze" (Bowden and Smith, 1994, p36). The results of the Pearson correlations for the present study, at least for the 9-10 and 11-12 year old subjects, confirm the findings with adults (Bowden and Smith, 1994) that scores from the first ten trials provide a good estimate of error scores to criterion. It should be stated that the Bowden and Smith article was published well after the current study had begun. In addition, Bowden and Smith's criticism of using the Austin Maze to detect frontal lobe damage rests on the assumption that the Maze is the only measure that would be used. This is in contrast to accepted clinical practice that a range of neuropsychological measures would be employed in assessment (Lezak, 1983).

## 9.11 Vertical Decalage

The issue of 'horizontal decalage', defined by Piaget (1976) as within-stage variability of performance, was discussed in Section 2.3. Piaget's research showed that there was significant variability of performance within each age group studied. According to Ginsburg and Opper (1988), Piaget did not, however, have a satisfactory explanation of individual differences in performances, particularly of variability that occurred within the same cognitive stage of thinking. Fischer and Bullock (1984) present evidence which clearly indicates that developmental variability in performance on Piagetian tasks is more the rule than the exception for children in the middle years of development (9-10 years of age). The present study highlights the incidence of what will be referred to here as **vertical decalage,** which by definition, is across-stage variability of performance. The notion of vertical decalage is one in which there is not a smooth and orderly reduction in variability in performance with age, but instead the age progression is characterised by significant variability. There is an unexpected instability in performance when one would expect a steady reduction with increasing age. As a group, the 9-10 year olds displayed an increased variability of performance, as evidenced by larger standard deviations, the occurrence of increased errors on several tests and a performance profile which deviated from a developmental curve, all of which is not evident in the performances of the 7-8 and 11-12 year old subjects.

For example, on the Fish Task, the 9-10 year olds scored the greatest number of perseverative errors, and obtained a standard deviation higher than the 7-8 year old group. See Table 11 and Figure 9 for the separate results. On the Balance Beam Task the 9-10 year olds obtained the largest standard deviation score for the Balance Beam Correct results (s.d. =9.65, 10.45 and 6.55 for Junior, Middle and Senior groups respectively). This variability was seen within a context of statistically significant increases in scores with increasing age (See Table 18 and Figure 20). On the Wisconsin Card Sorting Test, the 9-10 year olds appeared to be functioning quite differently to the other two age groups. They obtained the highest number of correct sets (2.9, 3.43 and 3.26 for 7-8, 9-10 and 11-12 year olds), the lowest number of perseverative errors (9.27, 7, 7.27), while scoring the largest variability in performance on perseverative errors (s.d.= 5.24, 5.82, 4.95). See Table 31 and Figures 31 and 35 for details of results.

The 9-10 year olds' performance on the Seals Task, in contrast to other test performances, displayed less variability than the other two age groups. On the Seals Total Score, while there is a statistically significant difference in scores at the Bonferroni adjusted level, the 9-10 year olds obtained the lowest standard deviation score. Standard deviations are 13.1, 11.65 and 13.25 for the 7-8, 9-10 and 11-12 year old subjects. A similar result occurred with Seals Monitoring scores where the standard deviations are 16.39, 12.385 and 18.305 for the 7-8, 9-10 and 11-12 year old groups. The 9-10 year olds clearly have less variability in their performance on the Seals Task when compared to the other two age groups. See Table 21 and Figures 22 and 25 for graphic representation of these results.

An inspection of the Austin Maze, AVLT and OLSAT scores reveals no significant variability in performance of the 9-10 year group. However, when compared to a similar sample of children tested on the AVLT (Forrester and Geffen, 1991), the present 9-10 year old subjects scored one standard deviation below on Learning and Retrieval indices. See Table 46 for details.

In summary, the present results indicate that the 9-10 year old group is functioning quite differently to the other age groups across a number of tests administered. This difference is observed in both standard executive function tests and the experimental psychological tasks. The 9-10 year olds' inconsistency in performance is characterised by larger variability as evidenced by higher standard deviations, by a decrement in performance on some tests, and even by enhanced performance compared to the oldest age group on some other tests. For example, on the WCST, the 9-10 year olds as a group obtained higher scores than the other two age groups. It is considered that the WCST result is further support for the inconsistency of performance of the Middle age group.

It is proposed that there are at least two plausible explanations for the general decrement in the performance of the 9-10 year old subjects used in this study. It may be that this is a biased group of subjects, somehow skewed as a result of sampling errors. The inconsistent performance of the 9-10 year group and the observation that somehow they were functioning quite differently to the other two age groups could be explained in terms of poor sampling techniques. An inspection of the sampling methods indicates that the Middle group was drawn from the same group of subjects as the other groups. The majority of subjects (88%) came from a *seriatim* sample of children at the one school. This proportion was similar to the 7-8 year group (87%), although varied from the 11-12 year group where 50% were drawn from the same school. The remaining subjects were drawn at random from three other government schools within the same geographic region which had been demonstrated to be comparable to the Sydney City population as determined by the 1991 Australian Bureau of Census. All children whose parents gave permission to be involved in the research were tested, and all subjects were tested by the same experimenter. In addition, it is contended that the sample size in the present study, with at least sixty subjects in each age group, is considered to be more than an adequate sized group for statistical purposes. Those outliers whose performance was significantly deviant from the main group were also eliminated from the sample.

A comparison of OLSAT test results for the 9-10 year old group with those of the OLSAT normative sample as outlined in the manual, indicates that there is no statistically significant difference in scores. The 9-10 year olds scored within the Low Average range of school ability, within one standard deviation of the normative sample (88.25, s.d. = 15.82, for OLSAT Total Score, compared to 100, s.d. = 16 for normative sample from OLSAT manual). It is argued that the 9-10 year old subjects in the present study are within acceptable limits of the OLSAT normative sample, suggesting that on a non-executive measure, the 9-10 year olds do not represent a skewed sample. If the premise that the present 9-10 year olds are a representative sample is accepted, then what explanation can be advanced for the variability and inconsistency of performance that gives rise to the vertical decalage phenomenon?

In Piagetian terms, it is contended that the 9-10 year olds are ‘caught’ between two stages of cognitive development: Concrete and Formal Operational Reasoning. As Late Concrete Operational Reasoners, this group is moving toward Early Formal Operational thought. The shift represents a significant cognitive growth (Flavell, 1985) characterised by a large-scale reorganisation of behaviours across many domains (Fischer and Bullock, 1984). The 9-10 year old is in a major transition period that will, according to Piaget, propel him from being an empirico-inductive reasoner to a hypothetico-deductive reasoner. The change is both dramatic and significant as it leads to the capacity to develop a new ability to generalise across concrete instances and to handle the complexities of tasks requiring hypothetical reasoning. Preadolescents, for example, can handle general definitions for a concept such as addition or a noun (Flavell, 1985), and can construct all possible combinations of four types of coloured blocks (Fischer and Bullock, 1984).

There is some evidence that the transition period is marked by a transient drop in the stability of behaviours across a sample of tasks (McCall, 1983). The variability in performance of this middle age group may be a phenomenon which has gone largely unnoticed in the literature. For example, in her study of 7-9 year old subjects performing the Austin Maze, Hume (1986) found that her 9 year olds incurred more errors than the 8 year old subjects. In addition, there was greater variability amongst the performance of the 9 year olds, when compared to the 8 year olds, with standard deviations of mean errors being 74.3, 46.4 and 66 for the 7, 8 and 9 year old groups, respectively. Hume suggested that this result may have been a function of her low sample size. There is also some evidence that other researchers have published test results which support vertical decalage for their 9-10 year old subjects, although have not commented on this phenomenon. It is understandable that a researcher, faced with a skewed result for this age group would offer an explanation in terms of sample bias, or some kind of aberration, as continuing improved progression in test performance is expected.

For example, a re-examination of the Rosselli and Ardilla (1993) study of 5-12 year old children's performance on the WCST indicates a larger variability in scores for their 9-10 year old subjects. This fact was not reported by the authors in their paper. A re-examination of their data on Correct Responses, shows standard deviations of 15.1, 11.6, 14.2 and 9.8 for 5-6, 7-8, 9-10 and 11-12 year old children, respectively. Note that the 9-10 year olds incurred a variability larger than the 7-8 and 11-12 year olds. On the category of Errors, the standard deviations are 23, 20.9, 22.3 and 19.1 for the three age groups. Once again, *post hoc* comparisons indicate that the 9-10 year old subjects had a larger variability in performance than the 7-8 and 11-12 year olds. Figures 54 and 55 below illustrate graphically how the 9-10 year old subjects are skewed in their standard deviation scores, compared to the other three age groups.

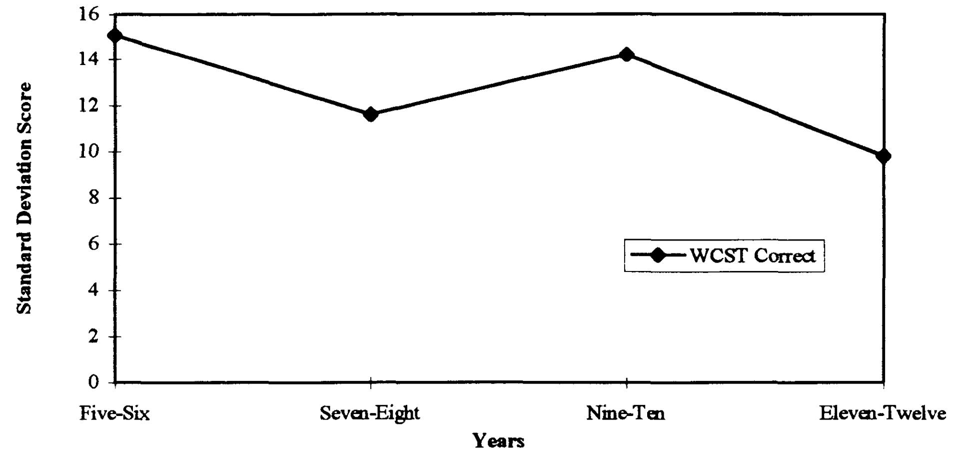

***Figure 54: Comparison of standard deviation scores on WCST correct responses across four age groups. From data reported by Rosselli and Ardilla (1993)***

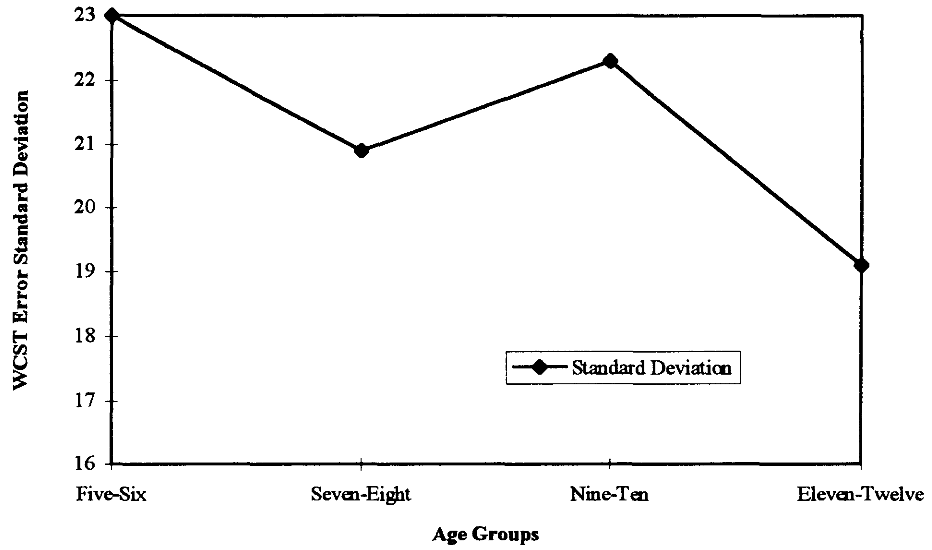

***Figure 55: Comparison of standard deviation scores across four age groups on WCST Error Responses. From data reported by Rosselli and Ardilla (1993)***

In a seminal study of performance of children at four age levels from 6 to 12 years, Passler, Isaac and Hynd (1985) examined verbal and nonverbal proactive and retroactive inhibition abilities. The authors argue that performance on these tasks is indicative of frontal lobe functioning and conclude that the greatest period of development appears to occur at the 6 and 8 year old levels. A re-examination of their reported scores indicates that performance on the retroactive inhibition task highlights variability in performance of both 10 year old boys and girls, a fact which was not reported by Passler *et al* (1985). For example, standard deviation data for normal girls (n=8) are reported as 4.7, 3.3, 5.2, and 3.6, for 6, 8, 10 and 12 year old subjects, respectively. The 10 year olds clearly had the largest variability in performance on this task and appear to be functioning quite differently to the other three age groups. Figure 56 below graphically portrays the rise in standard deviation for the 10 year old subjects.

***Figure 56: Comparison of standard deviation scores for a retroactive inhibition activity. From data reported by Passler, Isaac and Hynd (1985).***

The Geffen *et al* (1993) study of children's performance on the AVLT also highlights increased variability in the performance of the 9-10 year old group in comparison with other age ranges. For example, higher standard deviation scores are obtained for the 9-10 year old group on scores on the Distractor List B (Proactive Interference) and on Retention List A following the recall of the competing List B words (Retroactive Interference). The standard deviations reported in the authors' manual are 1.7, 2.3 and 1.4 for 7-8, 9-10 and 11-12 year old groups on the Distractor score, and 2.0, 2.2 and 1.4, for 7-8, 9-10 and 11-12 year old groups on Retention scores. This variability was not reported by Geffen *et al* (1993), although is consistent with the re-examination of the Passler *et al* (1985) data, where performance on a retroinhibition task produced higher standard deviation scores for the 10 year old subjects compared to the other age groups in the study. See Figure 56 above.

Is vertical decalage observable in particular tests or activities? An examination of the data indicates that this phenomenon is not present in four of the tests administered: OLSAT (Verbal and NonVerbal) , Austin Maze, AVLT and WCST Time Taken. In those tests, the performance of the 9-10 year olds is consistent with developmental improvement in that the raw scores for all of the measures taken increased in proportion to the 7-8 year olds. In addition, the standard deviation scores display a similar reduction in proportion to the 7-8 year old subjects. The first observation to be made is that the two OLSAT variables of Verbal and NonVerbal, the AVLT variable, as well as the WCST Time Taken, are the four variables which load significantly onto the Memory/Learning Factor. It is apparent that tests which appear to have Memory/Learning as an underlying factor, do not display vertical decalage in the performance of the 9-10 year old subjects. The two Austin Maze variables of Time Taken and Mean Errors, which also do not display vertical decalage, load onto the Executive Function Factor.

There does not appear to be a clear cut delineation, between the underlying factor structure, of those tests which display vertical decalage. There is evidence, however, that all the tests used in the current study, which had significant loadings on the Memory/Learning factor, do not display variability in performance for the 9-10 year old group. For this factor, scores are in the expected direction. It is contended that this result is consistent with the current CN Model. The argument has been maintained that those periods of 'cognitive disequilibrium', which occur prior to the child moving from one cognitive stage to a higher one, are manifest in activities which involve the interaction between active memory and planning processes. It is at these points in development where cognitive growth occurs as a result of internal conflict induced by a mismatch between the cognitive demands of an activity and the child's current store of problem-solving strategies. It is suggested that tasks which draw on executive function processes- the interaction between active memory and planning- would be those most sensitive to developmental changes occurring during the periods of cognitive growth.

The corollary would also hold. Non-executive tasks would be those least sensitive to developmental changes in the child. The results of the current study tend to confirm this assumption. The underlying Executive Function Factor is the only one of the three identified factors which displays a significant developmental improvement across the three age groups sampled in the current study. It has been argued that vertical decalage occurs during those same periods of cognitive growth, and is a direct result of internal cognitive conflict, or disequilibrium. It follows that vertical decalage would be most pronounced during performance on tasks which bring active memory and planning processes into focus. The present result that vertical decalage did not occur on tests which loaded onto the Memory/Learning Factor is therefore viewed as being consistent with the proposed model.

A sticking point in the preceding argument appears to be the fact that two Austin Maze variables, Time Taken and Mean Errors, both of which load onto the Executive Function Factor, also did not display evidence of vertical decalage. It would be anticipated that vertical decalage would be most noticeable during the performance on executive function tasks. This apparently is not the case for the Austin Maze test, and there does not appear to be an obvious explanation for this result. It may be another instance of unpredictability or instability in the performance of the 9-10 year old subjects. There may be some factor inherent in the Austin Maze which accounts for this aberration.

As indicated previously, Piaget (1976), does not pursue the notion of individual differences in performance and does not offer an explanation for horizontal decalage. The present results are consistent with a drop in the stability of behaviours and clearly demonstrate that the transition between Concrete Operational Reasoning and Formal Operational Reasoning is marked by both a decrement in and inconsistency of performance. It appears that as the 9-10 year old undergoes a qualitative change in underlying logical thinking processes, the child is often ‘between two posts’, sometimes using familiar Concrete Operational Reasoning strategies and sometimes struggling to employ Formal Operational Reasoning strategies. Is there a corollary for this explanation in neurological or physiological terms? What changes are occurring that may account for a decrement in or instability of performance at the 9-10 year old stage?

It has been previously established (See Section 2.3) that neurodevelopmentally, the age of 10 years represents a significant developmental milestone. From the age of 10, the levels of a number of brain “enzymes” begin to change significantly until they reach adult levels during adolescence. This finding applies to choline acetyltransferase, ChAT, (Court *at al*, 1993), and S-adenosylmethionine (SAMDC), (Morrison *et al*, 1993). ChAT is generally believed to be central to cognitive and memory processes (Dunnett, 1991). There is also a steady increase in glucose utilisation levels in the human cerebral cortex up to the age of 9 years, at which time there is a decline, with adult levels being reached by late adolescence (Gibson, 1991). In the papers which have examined levels of brain enzymes with increasing age, there has been no speculation of how this would affect cognitive function. For example, Gibson (1991, p41) does not offer an explanation for how the decline in glucose utilisation levels around the age of 9 years would affect cognitive function. It is therefore unclear to what extent a decline or increase in the level of brain enzymes around 9 to 10 years of age would impact on cognitive function. Welsh, Pennington, Ozonoff, Rouse and McCabe (1990, p1707), in a study examining executive functioning of children with early-treated Phenylketonuria, found support for their hypothesis "of a biochemical mechanism underlying executive functioning in PKU". Further research is required to substantiate the effect of neurochemistry on executive functioning.

The onset of puberty also provides a timetable which appears to impact significantly at 9-10 years of age. According to Shonkoff (1984), the normal onset of puberty for males begins at 9 years of age.

In summary, a significant finding of this study has been the discovery of a vertical decalage phenomenon in the performance of children around 9 to 10 years of age. This variability of performance is most noticeable on tests of problem-solving and executive functioning, and may be explainable in terms of both cognitive and physiological development. There are significant neurochemical and hormonal changes occurring within the human around the age of 9-10 years, as well as significant mental growth changes as the child moves from a logic of Concrete Operations to a logic of Formal Operational Reasoning at the same age. As a result of these significant changes, the child of 9-10 years is thrown into a state of Piagetian 'disequilibrium'. This 'disequilibrium' is a result of internal conflict caused by a mismatch between the cognitive demands of the situation and the strategies that the child has at his disposal. It is contended that at 9 to 10 years of age, more so than at other age levels in the developing child, the extent of Piagetian disequilibrium, normally associated with transition from one cognitive stage to the next, may be heightened by the significant changes that are also occurring in levels of hormones and brain enzymes. The impact, on the 9 to 10 year old child, of being in transition from Concrete Operational Reasoning to Formal Operational Reasoning, as well as being in transition from one level of hormone or brain enzyme, to another, would combine to create the level of variability, inconsistency, and decrement in performance that has been described as an outcome of this study.

It is at this point in development that the notion of discontinuity of cerebral maturation really comes into focus, (Hudspeth and Pribram, 1992; Thatcher, 1991). According to Thatcher (1991, p398), Piaget's description of cognitive development is characterised by an "integration with previous stages, a sort of spiral staircase in which contextual abstractions are re-examined each time the spiral sweeps around, forming successively higher levels of integration". Piaget's (1975) cognitive staging model asserts that there are periods of smooth and continuous growth while the child is in a state of equilibrium, only to be periodically interrupted by brief periods of rapid change, or disequilibrium, as the child moves from one cognitive stage to a higher one. Thatcher's (1991, p415) study, measuring EEG coherence, indicates that the period of Concrete Operational Reasoning is one age "that correlates to the age of maximum frontal lobe growth spurts". It is contended that the age span of 9-10 years is a period in development when discontinuity of cerebral and cognitive maturation is possibly at its highest level, and gives rise to the variability and inconsistency of performance that has been demonstrated in this study. There is a need to further explore this concept and this point is taken up in Section 9.18.

### 9.12 Discrimination of adult neuropsychological tests to developmental changes.

The results indicate that of all the tests administered, the Seals task produces the greatest sensitivity to developmental improvement across the three age groups, as well as being the only test which provides a clear separation between the brain-injured subjects and the matched normative group.

How sensitive were the adult neuropsychological measures to developmental change? Results of the WCST indicate a trend toward developmental improvement for the Perseverative Errors and Time Taken variable only. A statistically significant difference in performance is found for Time Taken between the Middle and Senior age groups and between the Junior and Senior groups, but not for the Junior and Middle age groups. There is a decrease in time taken to complete the task with increasing age. This result suggests a two-stage level of performance on the test, with the 7-8 and 9-10 year old group forming the first stage of maturation, and the 11-12 year old subjects forming the second developmental level. It appears that sensitivity to developmental changes on the variables of Perseverative Errors and Time Taken, applies only to discriminating the performance of the two maturational levels. The WCST, employed in the present study, is not sensitive to developmental changes in children below the age of 9-10 years.

In addition, it has been previously reported in Section 5, that the number of Correct Sets increases with age between 7-8 and 9-10 years, and then declines for the 11-12 year age group. This unexpected result appears to add further weight to the lack of sensitivity of the WCST in a population of children.

Previous studies with children, employing the WCST, have found at least three stages of maturation in solving the task. These have been at age 6, age 10 and during adolescence (Welsh *et al*, 1991, Passler *et al*, 1985, Chelune *at al*, 1986). Levin *et al* (1991), for example, in a study investigating developmental changes in performance on the WCST with a group of children in the 7-12 year age range, found evidence for significant improvement in scores on categories obtained. "The largest developmental shift occurred between the 7-8 and the 9-12 year age groups" (Levin *et al*, 1991, p385).

The current study did not replicate these previous findings of developmental sensitivity with age. It is contended that the different methodology and sampling employed in the present study may have accounted for the difference in results. In the current study, the WCST is presented on a microcomputer using the Nelson (1976) and Lau and Perdices (1990) adaptations of the task. No previous studies that have been reported have employed this presentation of the test. In addition, the sampling used in this study, was carefully drawn from populations which are considered to be representative of the Sydney City population, and approximately sixty subjects were drawn from each age group. By contrast, the Levin *et al* (1991) study, for example, involves only eighteen children in each age group, and chose subjects on the basis of a response to a newspaper advertisement.

On the AVLT, the Learning variable, which is the total number of words correctly recalled over the five learning trials, and the Retrieval variable, which is the proportion of Trial 8 words correctly recalled following a thirty minute delay, produced a developmental improvement between the Junior and Senior groups and between the Middle and Senior groups, but not between the Junior and Middle age groups. As previously raised in Section 5, the result suggests that, on these two variables, performance between the 7-8 and 9-10 year old children is indistinguishable, although significantly improves for the 11-12 year old subjects. The results suggest a two stage level of maturation for performance of the AVLT. The sensitivity of Retrieval efficiency and the Learning variable to developmental changes, is consistent with previous findings by Forrester and Geffen (1991).

As in the WCST, the results of the AVLT suggest that the test is not statistically sensitive to changes in performance for those children below the ages of 9-10 years.

On the Austin Maze Test, the error variable appears to be sensitive to developmental changes in performance. Mean Errors/Trial and Total Errors display developmental improvement with age. There is a statistically significant difference in scores for both error variables between the Junior and Senior and the Middle and Senior age groups, but not between the Junior and Middle age groups. A *post hoc* inspection of scores indicates a trend toward reduced errors with increasing age for each age group. Figure 40, for example, depicts the gradient of decline in errors for the three age groups. Results of statistical analysis, then, confirm a two-stage process for error scores on the Austin Maze. Children in the 7-8 and 9-10 year age groups form the first stage of development, and children in the 11-12 year age group comprise the second maturational stage. It has already been stated that adult performance on the Austin Maze occurs at 12 years of age (Walsh, 1991; Hume, 1986).

It is contended that the Austin Maze displays sensitivity to improvement in developmental age, although, from the current analysis, this sensitivity is not apparent in children below the age of 9-10 years.

In summary, the evidence presented in the current study suggests that for the WCST, the AVLT and the Austin Maze, three traditional neuropsychological tests, there is a lack of sensitivity to developmental changes in performance for children below the age of 9-10 years. The three tests produce a two-stage level of maturation of performance. The performance of the 7-8 year olds is statistically indistinguishable from that of the 9-10 year old group.

It is suggested that the lack of developmental sensitivity in the three adult neuropsychological measures, confirms the predictions made from the CN Model. From Table 2, it was predicted that the three measures did not contain a graded level of difficulty, were not based on developmental theory or research and consequently would not be sensitive to cognitive shifts in performance across the three age ranges studied in the present research.

### 9.13 Brain-Injured Sample

Medical records of the twenty brain-injured children were used to confirm neurological insult to the brain. Although only twelve of the twenty subjects had abnormal CT scans, all twenty had experienced a trauma of the acceleration/deceleration type, and all had been hospitalised for a period of ten days as a consequence of their injuries. Mean period of time since the brain injury was two years ten months.

The performance of the brain-injured sample was compared with the matched normative group. Both groups were matched for age, school ability and performance on the AVLT. In addition, all children were attending mainstream classes at the time of assessment. The results indicate that the Seals task is the only variable which provides an unequivocal separation between the two groups. The first implication of this outcome is that a heterogenous group of children with brain injuries performs significantly lower on a test of experimental executive function than a matched healthy group of subjects. It has been established that only five of the brain-injured group had confirmed frontal lesions or frontal connection damage (thalamus, internal capsule). The remaining subjects experienced more posterior damage. However, it is not the intention of this study to isolate particular areas of cortical damage nor to demonstrate a relationship between brain structural damage and cognitive function. The medical records were used simply to confirm the presence of brain damage for inclusion of the clinical sample into a category of "brain damage". However, having stated this caveat, the results suggest that lower performance on measures of executive function are observable in a sample of children with a range of brain injuries, including those with no apparent frontal signs. Given the rich connections of the prefrontal cortex with cortical and subcortical structures (Pandya and Barnes, 1987), the results, nonetheless are not surprising. Stuss (1992) prefers the term "frontal system" emphasising that the brain is an integrated functioning unit with the prefrontal cortex as the executive.

A *post hoc* inspection of the mean scores of the four experimental developmental measures, and Figure 50 depicting the results, indicates that the brain-injured sample performed more like the younger 7-8 year old group of subjects than the matched 9-10 year old normative sample. This result suggests that the 9-10 year old brain-injured subjects, as a group, are functioning more like Early Concrete Operational Reasoners, than Late Concrete Operational as one might expect from their chronological ages. Future studies should examine this issue in more detail to discover the cognitive correlates which are affected. The difference in performance is seen within the context of comparable performance on a measure of school ability.

Early Concrete Operational Reasoners are prone to centre their attention on a single highly salient task element, to base their judgments on perceived appearances rather than inferences and thinking tends to be irreversible (Flavell, 1985). The child's problem solving approaches are at an early empirico-inductive stage where propositions are considered singly, that is, in isolation from one another (Ginsburg and Opper, 1988). If it is accepted that the healthy 7-8 year olds and the 9-10 year old brain-injured group are both functioning at this early empirico-inductive level, then how is this result interpretable within a Piagetian framework in general and the CN Model in particular? Piaget did not elaborate on the implication of his stage model of cognitive development other than for normal, healthy children (Piaget, 1976). However, the performance of the brain-injured group suggests that there is a decrement in the development of those mental structures leading to the full attainment of executive functioning. Within the CN Model, it is proposed that the current results suggest there is a delay in the development of Formal Operational Reasoning. Further research is required to confirm the developmental timetable for the maturation of executive functions in a brain-injured paediatric population. Testing of 11-12 and 13-14 year old children with brain injuries and comparing their performances with the normative sample, as well as a sample of adults, would provide insight into charting the development of executive function skills in a population of brain-injured children. Would the 11-12 year old children with brain injuries, for example, function more like a healthy group of 9-10 year olds? If so, then this would suggest a learning gradient similar to a healthy population. It has been suggested that injury to the prefrontal lobes of children would result in significant failure in mental development (Stuss, 1992). In fact, there is a view that if damage occurs to the brain prior to frontal lobe maturation, late onset difficulties associated with reasoning and problem solving may occur (Anderson, V.,1992).

### 9.14 Canonical Correlation

The Canonical correlation, between the brain-injured and matched normative samples is 0.518, which is a significant result. The separate independent variable canonical loadings confirm that, of the independent variables, Seals accounts for the majority of the variance on the group variable with a correlation of 0.853 with the canonical value. On the other hand, other tests had relatively low correlation scores, with the next largest correlations being WCST Correct with a score of 0.407 and AVLT with a correlation score 0.269. Seals also had a coefficient of 0.882, again confirming the dominant role it played in separating the acquired brain-injured and normative groups.

The two Austin Maze measures, the AVLT score and OLSAT score have moderately low canonical loadings suggesting they add little to the variance in explaining the discriminant function.

### 9.15 Problem-solving strategies of brain-injured sample

An analysis of the problem-solving strategies of the brain-injured sample was taken from the same error patterns observed on Trials 1 and 2 of the 3-ball and 4-ball Seals tasks. For both problem types, the majority of errors occur on Move #1 (95% for the 3-ball and 50% for the 4-ball), with Moves #3 and #5 incurring the largest percentage of errors for Trial 2 errors. It appears that the same pattern of errors is observed in both normative and brain-injured samples. This result is consistent with previous findings by Levin, Goldstein, Williams and Eisenberg (1991) who found no difference between frontal, non-frontal and a healthy control group on error patterns on the first trial of each Tower of Hanoi problem.

In addition, samples of subjects' verbalisations during performance on both 3 and 4-ball problems confirm that the vast majority of utterances occur during those critical key error moves. This result is also consistent with findings from the normative sample of subjects. Incidence of self-corrections for this group provides some interesting statistics. Five subjects self corrected their responses, one during the 3-ball problem and four during the 4-ball problem. Three of the five self-corrections occurred at key error points. This result differs from the matched normative group, where all nine self-corrections occurred at the key choice or error points. Proportionately, the incidence of self-correction for the brain-injured group is greater than that for the normative sample (25% of brain injured subjects self-corrected compared with 15% of the matched normative group). However, the difference between the two groups may be explained by sampling factors. The brain-injured group contained only twenty subjects, some of whom were girls, compared with sixty subjects in the normative group which contained only boys carefully drawn from a population which represented the Sydney City region. These differences in sampling may have created a bias which affected the outcome.

### 9.16 Discriminant Power of Paediatric Versus Adult Measures of Executive Function

Given the magnitude of the Seals test statistic, then how significant a discriminator is it on its own? The canonical value of 0.518 sets the upper limit for any correlation that separates the two groups. A Pearson correlation matrix involving the ten variables reported in the canonical analysis as well as a group variable (1=normative sample, n=60 and 2=brain injured sample, n=20) was conducted between the normative and clinical samples using Systat. The correlation between Seals and the Group variable is r= 0.459, which approaches the canonical value, suggesting that on its own, the

Seals task is a good predictor of brain-injury involving children in the 9-10 year age range. The nearest correlation is WCST Correct with a correlation of r=0.239.

The evidence suggests that the Seals task, as a measure of executive function, has good construct validity (Borg and Gall, 1989) in that it clearly demonstrates developmental improvement in scores of a healthy population and it significantly differentiates the scores of an acquired brain-injured sample from a matched normative group of children. In addition, the Seals task was demonstrated to have a high level of specificity, with a Negative Predictive Power of 93%, suggesting that the Seals task would be recommended as a second-stage diagnostic tool in the detection of brain damage in children.

### 9-17 Limitations of Study

The main objective of this study was to examine the developmental performance of healthy children, according to Piagetian principles, on experimental developmental measures of executive function. The results indicate that the Seals task displays a statistically significant developmental improvement across the three age groups used. However, a measure of internal reliability was not obtained for this task, because not all subjects were administered all trials of each problem type. To overcome this problem, a test-retest paradigm could have been employed, where a smaller sample of subjects was administered the Seals on two separate occasions. It is argued that a practice effect may have confounded the validity of test-retest in this instant. However, a measure of test reliability is necessary to determine the error of measurement of a particular test, and to demonstrate that subjects tested over time will yield similar scores (Borg and Gall, 1989).

The sample size of at least sixty subjects in each of the three age ranges is considered to be within acceptable limits. However, there is an issue in relation to sampling procedures which may limit the generalizability of the conclusions that were reached. The demographic data, outlined on Table 9, indicate that the level of tertiary qualifications for the Sydney City normative sample was double that of the current sample, while the proportion of professional families was also lower for the present sample. However, the difference between the two samples was small and unlikely to have had a significant impact on the results.

It is also noted that the results of the OLSAT indicate that the 7-8 and 9-10 year old groups both performed on the low side of average, compared to a normative sample. Their mean Total

Scores were 88.24 and 88.25 respectively (compared with a mean standard score of 100). These results suggest a problem with sample bias with these two age groups. For example, the results of the AVLT appear to highlight significantly poorer scores for the current 9-10 year old group when compared with those of Geffen (1991). These lower scores on the AVLT and the OLSAT for the 9-10 year old group may be an indication that this particular group of subjects would not be a representative sample. The small sample size used by Geffen (1991) must be treated with caution. However, the findings of Geffen (1991) are in line with those of Anderson *et al* (1994) who employed a representative sample of 360 children.

There are also limitations to cross-sectional design and a longitudinal study would ideally be the preferred choice. However, the time constraints on a PhD thesis would make a longitudinal study impractical, given that there is a time span of more than five years across the ages studied in this project. In addition, it is considered that the range, from 7 years to 12 years of age, is reasonably confined and does not represent too large an age spread, suggesting that this cross-sectional design would not be problematic in terms of social and cultural influences that affect cross-sectional design with a large age spread.

In addition, the size and range of subjects in the brain-injured sample did not match the same standards. The experimenter matched the middle, 9-10 year old group, with a brain-injured group of subjects of the same age. Limitations are placed on the availability of subjects that matched the age and criterion of 9-10 years and attending a mainstream class at school. Twenty of the twenty-one possible subjects whose parents gave permission to be involved in the project and who satisfied the two stated criteria, are included in the sample. The one subject who was not included, lived some four hundred kilometres from Sydney. It is recognised that particular clinical samples often restrict the representative nature of these subjects. In the present study, however, nineteen of the twenty brain-injured subjects came from the same geographical area as the normative sample.

It could be argued that, by comparison with the normative sample, the brain-injured group is small (n=60, compared to n=20). It could also be argued that brain-injured samples for 7-8 and 11-12 year old subjects could also have been included in the study. The logic of this criticism, however, does not necessarily rest there, for one could also ask why 5-6 or 13-14 year old subjects are not included in the study. The age ranges were selected on the basis of both neurological and cognitive data identifying the three age groups as periods where rapid growth occurred in development. It is acknowledged that a possible limitation in the study is the choice of 9-10 years for comparison with the brain-injured subjects. One can only speculate whether the unexpected performance of the normative sample of 9-10 year old subjects had any significant influence on comparison between the two groups. As discussed previously, the 9-10 year old normative subjects displayed an inconsistency of performance that was marked mostly by a decrement in and large variability of scores across a number of executive function tasks. In other words, the 9-10 year olds did not appear to be an homogenous group. Arguments were advanced explaining this instability in performance in terms of a transitional period of cognitive development, neurophysiological changes that occur at 9-10 years, in addition to hormonal influences signalling the onset of puberty. These same physiological changes also apply to the matched brain-injured subjects. With hindsight, the greater variability of the 9-10 year old group may have compromised the size of the differences between the brain-injured and matched normative group. A comparison of 7-8 or 11-12 year old brain-injured subjects with matched normative groups may have avoided the possible contamination influences of a wide variability in performance.

It is acknowledged that following the Pilot Study, a second study might have been used to test the proposed modifications to the experimental developmental tasks. However, a second study may not have indicated significant changes to be made to the experimental tasks. Pilot studies conventionally contain low sample sizes and the resultant loss of statistical power may not have produced results that were reliably identified. The results of testing indicate that the Piggy Bank task, in particular, has low reliability and did not discriminate significantly between the three age groups. Modifications to this test have been proposed following the Normative Study. Perhaps a second Pilot study may have identified areas of weakness in the test which could have been addressed in the Normative Study. Of the other experimental developmental tests, the Seals task provides a clear separation between the three age groups and did not warrant further modification. Modifications to the Balance Beam and the Fish Task are recommended.

## 9-18 Directions for Future Research

There are a number of issues which are suggested from the present study that require further follow-up. The notion of continuing with further test development for the experimental developmental tests is one which should be pursued. It has been stated that results of analysis for the Fish and Piggy Bank tasks suggest specific changes to these tests which should be implemented and trialled with a further sample of children. This point has been raised in the previous section. For example, further development of the Fish task should include a component of reaction time which is recorded by the computer. The reaction time would then provide information on subjects' responses to the Target or Background fish to determine whether there are identifiable patterns of long or short reaction times to either stimulus. As raised in the Discussion section, it may be 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. A measurement of reaction time may confirm this hypothesis.

It is recommended that the experimental developmental tests be administered to a wider sample of children with brain injury than was used in this study. In the first instance, there is a need to gather data from a larger sample of 9-10 year old children with brain injuries. Do the tests which load onto the Executive Function Factor have the same degree of discriminatory power with 7-8 and 11-12 year old children with brain injuries as they did with a sample of 9-10 year old children with brain injuries? It is proposed to administer the revised tests to a sample of at least twenty children with brain injuries in the 7-8 and 11-12 year age group. It is proposed that these children would be attending mainstream classes, as did the group in the current study. The analysis, using a wider range of children with brain-injury, would enable confirmation of the results that are obtained with the current 9-10 year old group.

There is a need to obtain a measure of reliability of the Seals task and it is recommended that a test-retest correlation coefficient be obtained following a delay of three months between testing. The reliability index will determine the error of measurement (Borg and Gall, 1989) and will provide critical information on the robustness of the Seals test.

In addition, the results of the Balance Beam test, detailed in Section 9.2, indicate that the test in its present form does not support the CN Model. It is contended that one possible reason for the test's poor discrimination between a group of brain-injured children and a normal group, is its concept of active memory. As a test of this assertion, it is proposed to redevelop the Balance Beam test so that subjects do not receive feedback on their performance. This adjustment to the test structure would bring it closer to alignment to the internal structure of the Seals test, where subjects do not receive such feedback. The same items would be presented, subjects would still make a choice between one of three possible solutions (left side drops down, right side drops down, the beam balances), although the beam does not move to indicate whether the subjects made correct choices.

It is proposed to redesign the test and to maintain the same research design. The Balance Beam would be administered to a group of normal children in the 7-12 year age group, as well as a group of brain-injured children.

The concept of vertical decalage is one which is raised in this study as a possible explanation for the significant variability in performance of the 9-10 year old subjects. There is a need to replicate the phenomenon using a large representative sample of subjects. It may be that a meta-analysis would lend itself to this type of investigation where a large number of existing studies involving a range of children performing problem-solving type activities is studied. Some evidence from previous studies was presented, attesting to large variability in performance for this group of children. However, there is a need to systematically confirm the presence of the vertical decalage phenomenon.

The point was made in the Foreword, that one of the underlying motivations for the development of paediatric measures of executive functioning, is to assist in the implementation of cognitive rehabilitation programs for children. There is a view that theoretically-driven neuropsychological and behavioural assessments significantly influence the efficacy of rehabilitation programs (Wilson, 1991). It is proposed that the experimental developmental tests developed in this study be examined with a view to their usefulness in the design of cognitive rehabilitation programs. The neuropsychological rehabilitation literature does report the use of problem-solving tasks in the pre and post assessment of patients undergoing a period of intervention specifically designed to remediate problem-solving deficits (Von Cramon, Matthes-von Cramon and Mai, 1991; Von Cramon and Matthes-von Cramon, 1992). On the basis of the current results, it is suggested that the Seals task may be a useful test of executive function with children that could be employed as a pre and post-measure during an intervention study.

An advantage of using the Seals task, over more traditional adult neuropsychological measures, is that it lends itself to strategic analysis. An inspection of subjects' moves, during the critical key error points, would provide insight into the types of problem-solving strategies employed by subjects. At a functional level, this type of information would be of interest to those who are working with the subject in a rehabilitation setting. Once the subject's strategies were identified, they could become the basis for an intervention program. Post-testing, using the Seals task, would then be a direct test of the efficacy of the particular intervention.

In an unpublished study, funded by the NSW Motor Accident Authority, Bogan (1994), employed the Seals task as a pre and post measure for a problem-solving training program involving adolescents who had sustained an acquired brain injury as a result of a motor vehicle accident. The Seals task identified problem-solving deficits which became the focus for a twelve-session individual training program. Compared to a matched Control group who received no intervention, the three Intervention group subjects displayed improvement in teacher-observed adaptive behaviours. On the other hand, teachers rated all the Control subjects as actually regressing in adaptive behaviours in the classroom. All teachers were blind to the students' involvement in the program. In addition, two of the three Intervention subjects significantly improved their performance on the Seals task following the intervention period.

It is contended that, as well as having clinical utility in identifying executive function deficits in a paediatric population, the Seals task does represent a measure which may also have utility in a rehabilitation setting. Further research is required to support this contention.